MENDELSON

INTRODUCTION TO TOPOLOGY

$(X, \mathcal{T})$

BLACKIE

INTRODUCTION TO TOPOLOGY

MENDELSON

BLACKIE

Topology is the study of continuity and underlies every subject in mathematics and science that uses continuity. Until recently it was seldom taught in universities. Now it has an established place in the mathematics departments of most universities where the elementary principles are introduced during the first year.

Introduction to Topology has been well received in America, where it was first published. In his Preface to this British edition, Dr. E. C. Zeeman, Lecturer in Mathematics at Cambridge University, points out the book's two outstanding features. First, it is written from a geometrical point of view, which is important because today the habit of geometrical thinking influences even the most abstract mathematical subjects. Second, it contains all the essential material and no more. It starts from nothing and gently introduces the reader to the elementary principles. There are sufficient exercises and examples to make it an ideal introduction at any level, be it school, university or scientific laboratory.

35s net

Printed in Great Britain

Introduction to Topology

INTRODUCTION TO

TOPOLOGY

Bert Mendelson

ASSOCIATE PROFESSOR OF MATHEMATICS, SMITH COLLEGE
MASSACHUSETTS, U.S.A.

With a Preface to the British Edition by
E. C. Zeeman, M.A., Ph.D.
FELLOW OF GONVILLE AND CAIUS COLLEGE, CAMBRIDGE

Blackie & Son Limited
LONDON GLASGOW

British Edition first published in 1963

BLACKIE & SON LIMITED
5 FITZHARDINGE STREET
PORTMAN SQUARE
LONDON · W.1
BISHOPBRIGGS, GLASGOW

BLACKIE & SON
(INDIA) LIMITED
103-5 FORT STREET
BOMBAY

Printed and bound by Blackie & Son Ltd. at the Villafield Press, Glasgow

Preface

BY DR. E. C. ZEEMAN

Topology and algebra are the two main pillars upon which modern mathematics is built. The student who has grasped the elementary principles of these two subjects discovers a simplicity and coherence running through the whole of mathematics. It is the elementary principles that are important rather than the technical details, and the idea of this book is to start from nothing and gently introduce the reader to the elementary principles of topology.

Topology is the study of continuity, and underlies every subject in mathematics and science that uses continuity. In particular topology underlies calculus, and logically ought to be taught before calculus. Often the schoolboy who cannot understand calculus when he first meets it is a better mathematician than the one who swallows it. The limiting process of calculus is a sophisticated technical process, inclined to impart a flavour of cheating to the boy who has previously been happy with the precision of arithmetic, algebra and geometry. This is disastrous because haziness of understanding produces fear of mathematics. Such a boy ought to be given this book. If he is mature enough to start calculus then he is mature enough to understand topology. By grasping the topological principles underlying calculus he will recapture that feeling of precision.

Ten years ago topology was taught, if at all, only in the third year at English universities, but today in recognition of the logical place of the subject, most universities introduce the elementary principles in the first year. This book covers exactly that first year material. In ten years' time it may well be taught in schools.

Compared with previous introductions to the subject, this book has two outstanding features. First, it is written from a geometrical point of view: there are plenty of diagrams and the reader is encouraged to draw his own, and to think geometrically. This is important because today the habit of geometrical thinking influences even the most abstract mathematical subjects. Also in higher dimensions, the most geometrical subject in mathematics is no longer geometry but algebraic topology, which is the sequel to this book.

Secondly, this book contains all the essential material and no more; there is room to relax, and there are sufficient exercises and examples to make it an ideal introduction at any level, be it school, university or scientific laboratory.

The first chapter consists of the usual discussion of set theory. The concept of a diagram consisting of sets and functions has been introduced at the same time. The concepts of equivalence relation and countability have been reserved for mention later, in Chapters IV and V respectively, where they make a natural appearance in connection with other topics.

The second chapter is a discussion of metric spaces, where the topological terms *open set, neighbourhood*, etc., have been carefully introduced. Particular attention is paid to various distance functions which may be defined on Euclidean n-space and which lead to the ordinary topology.

In the third chapter, topological space is introduced as a generalization of metric space. A great deal of attention has been paid to alternative procedures for the creation of a topological space, using neighbourhoods, etc., in the hope that this seemingly trivial, but subtle, point may be clarified. Since topological space is a generalization of metric space, it is hoped that the reader will observe the similarity, or perhaps redundancy, in the presentation of these two topics.

Chapters IV and V are devoted to a discussion of the two most important topological properties, connectedness and compactness. As applications the reader is introduced to a little algebraic topology. In Chapter IV to explain simple connectedness the concepts of homotopy and the fundamental group are described, except that the group structure is omitted because the reader is not presumed to know any group theory. Chapter V is concluded with a discussion of two-dimensional closed surfaces.

May, 1963 E. C. ZEEMAN

Contents

Introduction to Topology

Theory of Sets

I

1 Introduction

As in any other branch of mathematics today, topology consists of the study of collections of objects that possess a mathematical structure. This remark should not be construed as an attempt to define *mathematics*, especially since the phrase "mathematical structure" is itself a vague term. We may, however, illustrate this point by two important examples.

The set of *positive integers* or *natural numbers* is a collection of objects N on which there is defined a function s, called the *successor function*, satisfying the conditions:

1. For each object x in N, there is one and only one object y in N such that $y = s(x)$;
2. Given objects x and y in S such that $s(x) = s(y)$, then $x = y$;
3. There is one and only one object in N, denoted by 1,

which is not the successor of an object in N, i.e., $1 \neq s(x)$ for each x in N;

4. Given a collection T of objects in N such that 1 is in T and for each x in T, $s(x)$ is also in T, then $T = N$.

The four conditions enumerated above are referred to as *Peano's axioms for the natural numbers.* The fourth condition is called *the principle of mathematical induction.* One defines addition of natural numbers in such a manner that $s(x) = x + 1$, for each x in N, which explains the use of the word "successor" for the function s. What is significant at the moment is the conception of the natural numbers as constituting a certain collection of objects N with an additional mathematical structure, namely the function s.

We shall use the system of real numbers as a second example of the fact that the type of entity that one studies in mathematics is a collection of objects with a certain mathematical structure. This explanation will require several preliminary definitions.

A *commutative field* is a collection of objects F and two functions that associate to each pair a, b of objects from F an element $a + b$ of F, called their sum, and an element $a\ b$ of F, called their product, respectively, satisfying the conditions:

1. For each a, b in F, $a + b = b + a$;
2. For each a, b, c in F, $a + (b + c) = (a + b) + c$;
3. There is a unique object in F, denoted by 0, such that $a + 0 = 0 + a = a$ for each a in F;
4. For each a in F, there is a unique object a' in F such that $a + a' = a' + a = 0$;
5. For each a, b in F, $a \cdot b = b \cdot a$;
6. For each a, b, c in F, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
7. There is a unique object in F, different from 0, denoted by 1, such that $a \cdot 1 = 1 \cdot a = a$ for each a in F;
8. For each a in F, if a is different from 0, there is a unique object a^* in F such that $a \cdot a^* = a^* \cdot a = 1$;
9. For each a, b, c in F, $a \cdot (b + c) = a \cdot b + a \cdot c$.

A commutative field F is thus a set of objects and an addition and multiplication that satisfies rules analogous to the rules of addition and multiplication of real numbers.

A field is called *linearly ordered* if it has additional structure, namely a relation "$<$" which satisfies the properties of "less than" as used in the real number system. Precisely, a field F is called a *linearly ordered* field if there is a relation "$<$" among certain ordered pairs of objects of F satisfying the conditions:

1. For each pair of objects x, y in F, one and only one of the three statements, $x < y$, $x = y$, $y < x$, is true;
2. For each object z in F, $x < y$ implies $x + z < y + z$;
3. For each object z in F such that $0 < z$, $x < y$ implies $x \cdot z < y \cdot z$.

Let T be a subcollection of objects from a linearly ordered field F. An object b in F is called an *upper bound* of T if for each x in T, either $x < b$ or $x = b$. An object a in F is called a *least upper bound* of T, if a is an upper bound of T and if $a < b$, where b is any other upper bound of T.

As a final definition before describing the system of real numbers, a linearly ordered field F is called *complete* if every non-empty subcollection T of F that has an upper bound also has a least upper bound. We can now state that the real number system is a collection R of objects together with operations of addition and multiplication and a relation "$<$" such that the collection R, together with this structure, is a complete, linearly ordered, commutative field.

[The use of the definite article, "*the* real number system is . . . ," should not be construed as asserting that there is only one real number system, but it is implicitly asserted that the conditions imposed on the collection R are categorical; that is, that any two instances of the real number system are indistinguishable, apart from the names or notation used to denote the objects.]

Thus we see that some of the better-known mathematical

objects of study are describable as collections of objects together with certain specified structures. We shall describe a *topological space* in the same terms, although an appreciation of the utility of this concept can only come later. A topological space is a collection of objects (these objects usually being referred to as points), and a structure that endows this collection of points with some coherence, in the sense that we may speak of nearby points or points that in some sense are close together. This structure can be prescribed by means of a collection of subcollections of points called *open sets*. As we shall see, the major use of the concept of a topological space is that it provides us with an exact, yet exceedingly general setting for discussions that involve the concept of continuity.

By now the point should have been made that topology, as well as other branches of mathematics, is concerned with the study of collections of objects with certain prescribed structures. We therefore begin the study of topology by first studying collections of objects, or, as we shall call them, *sets*.

2 Sets and Subsets

We shall assume that the terms "object," "set," and the relation "is a member of" are familiar concepts. We shall be concerned with using these concepts in a manner that is in agreement with the ordinary usage of these terms.

If an object A belongs to a set S we shall write $A \in S$ (read, "A in S"). If an object A does not belong to a set S we shall write $A \notin S$ (read, "A not in S"). If $A_1, \ldots, A_n$ are objects, the set consisting of precisely these objects will be written

$$\{A_1, \ldots, A_n\}.$$

For purposes of logical precision it is often necessary to distinguish the set $\{A\}$, consisting of precisely one object A from the object A itself. Thus

$$A \in \{A\}$$

is a true statement, whereas

$$A = \{A\}$$

is a false statement. It is also necessary that there be a set that has no members, the so-called *null* or *empty* set. The symbol for this set is $\emptyset$ (a letter in the Swedish alphabet).

Let A and B be sets. If for each object $x \in A$, it is true that $x \in B$, we say that A is a *subset* of B. In this event, we shall also say that A *is contained in* B, which we write

$$A \subset B,$$

or that B *contains* A, which we write

$$B \supset A.$$

In accordance with the definition of subset, a set A is always a subset of itself. It is also true that the empty set is a subset of A. These two subsets, A and $\emptyset$, of A are called *improper* subsets, whereas any other subset is called a *proper* subset.

There are certain subsets of the real numbers that are frequently considered in calculus. For each pair of real numbers a, b with $a < b$, the set of all real numbers x such that $a \leqq x \leqq b$ is called the *closed interval* from a to b and is denoted by $[a, b]$. Similarly, the set of all real numbers x such that $a < x < b$ is called the *open interval* from a to b and is denoted by (a, b). We thus have $(a, b) \subset [a, b] \subset R$, where R is the set of real numbers.

Two sets are identical if they have precisely the same

members. Thus, if A and B are sets, $A = B$ if and only if* both $A \subset B$ and $B \subset A$. Frequent use is made of this fact in proving the equality of two sets.

Sets may themselves be objects belonging to other sets. For example, $\{\{1, 3, 5, 7\}, \{2, 4, 6\}\}$ is a set to which there belong two objects, these two objects being the set of odd positive integers less than 8 and the set of even positive integers less than 8. If A is any set, there is available as objects with which to constitute a new set the collection of subsets of A. In particular, for each set A, there is a set we denote by 2^A whose members are the subsets of A. Thus, for each set A, we have $B \in 2^A$ if and only if $B \subset A$.

Exercises

1. Determine whether each of the following statements is true or false:

(a) For each set A, $\emptyset \in A$.

(b) For each set A, $\emptyset \subset A$.

(c) For each set A, $A \subset A$.

(d) For each set A, $A \in \{A\}$.

(e) For each set A, $A \in 2^A$.

(f) For each set A, $A \subset 2^A$.

(g) For each set A, $\{A\} \subset 2^A$.

(h) $\emptyset \in \{\emptyset\}$.

(i) For each set A, $\emptyset \in 2^A$.

(j) For each set A, $\emptyset \subset 2^A$.

(k) There are no members of the set $\{\emptyset\}$.

(l) Let A and B be sets. If $A \subset B$, then $2^A \subset 2^B$.

* The compound statement "P if and only if Q," is the conjunction of the two statements "If P then Q" and, "If Q then P." A statement of the form "P if and only if Q" may also be phrased "If P then Q and conversely."

(m) There are two distinct objects that belong to the set $\{\emptyset, \{\emptyset\}\}$.

2. Let A, B, C be sets. Prove that if $A \subset B$ and $B \subset C$, then $A \subset C$.

3. Let $A_1, \ldots, A_n$ be sets. Prove that if $A_1 \subset A_2$, $A_2 \subset A_3, \ldots,$ $A_{n-1} \subset A_n$ and $A_n \subset A_1$, then $A_1 = A_2 = \ldots = A_n$.

4. Let A be a set to which there belong precisely n distinct objects. Prove that there are precisely 2^n distinct objects that belong to 2^A.

3 Set Operations: Union, Intersection, and Complement

If x is an object, A a set, and $x \in A$, we shall say that x is an *element, member,* or *point* of A. Let A and B be sets. The *intersection* of the sets A and B is the set whose members are those objects x such that $x \in A$ and $x \in B$. The intersection of A and B is denoted by

$$A \cap B$$

(read, "A intersect B"). The *union* of the sets A and B is the set whose members are those objects x such that x belongs to at least one of the two sets A, B; that is, either $x \in A$ or $x \in B$.* The union of A and B is denoted by

$$A \cup B$$

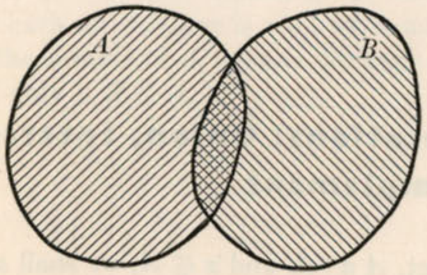
(read, "A union B").

The operations of set union and set intersection may be represented pictorially (by *Venn diagrams*). In Figure 1, let

* The logical connective "or" is used in mathematics (and also in logic) in the inclusive sense. Thus, a compound statement "P or Q" is true in each of the three cases: (1) P true, Q false; (2) P false, Q true; (3) P true, Q true, whereas "P or Q" is false only if both P and Q are false.

the elements of the set A be the points in the region shaded by lines running from the lower left-hand part of the page to the upper right-hand part of the page, and let the elements of the set B be the points in the region shaded by lines sloping in the opposite direction. Then the elements of $A \cup B$ are the points in a shaded region and the elements of $A \cap B$ are the points in a cross-hatched region.

Figure 1



Let $A \subset S$. The *complement* of A in S is the set of elements that belong to S but not to A. The complement of A in S is denoted by $C_S(A)$ or by $S - A$. The set S may be fixed throughout a given discussion, in which case the complement of A in S may simply be called the complement of A and denoted by $C(A)$. $C(A)$ is again a subset of S and one may take its complement. The complement of the complement of A is A; that is, $C(C(A)) = A$.

There are many formulas relating the set operations of intersection, union, and complementation. Frequent use is made of the following two formulas.

Theorem (DeMorgan's Laws) Let $A \subset S$, $B \subset S$. Then

(3.1) $$C(A \cup B) = C(A) \cap C(B),$$

(3.2) $$C(A \cap B) = C(A) \cup C(B).$$

Proof. Suppose $x \in C(A \cup B)$. Then $x \in S$ and $x \notin A \cup B$.

Thus, $x \notin A$ and $x \notin B$, or $x \in C(A)$ and $x \in C(B)$. Therefore $x \in C(A) \cap C(B)$ and, consequently,

$$C(A \cup B) \subset C(A) \cap C(B).$$

Conversely, suppose $x \in C(A) \cap C(B)$. Then $x \in S$ and $x \in C(A)$ and $x \in C(B)$. Thus, $x \notin A$ and $x \notin B$, and therefore $x \notin A \cup B$. It follows that $x \in C(A \cup B)$ and, consequently,

$$C(A) \cap C(B) \subset C(A \cup B).$$

We have thus shown that

$$C(A) \cap C(B) = C(A \cup B).$$

One may prove Formula 3.2 in much the same manner as 3.1 was proved. A shorter proof is obtained if we apply 3.1 to the two subsets $C(A)$ and $C(B)$ of S, thus

$$C(C(A) \cup C(B)) = C(C(A)) \cap C(C(B)) = A \cap B.$$

Taking complements again, we have

$$C(A) \cup C(B) = C(C(C(A) \cup C(B))) = C(A \cap B).$$

Exercises

1. Let $A \subset S$, $B \subset S$. Prove the following:

(a) $\emptyset = C(S)$.
(b) $S = C(\emptyset)$.
(c) $A \cap C(A) = \emptyset$.
(d) $A \cup C(A) = S$.
(e) $A \cap A = A$.
(f) $A \cup A = A$.
(g) $A \cup S = S$.
(h) $A \cap S = A$.
(i) $A \cup \emptyset = A$.
(j) $A \cap \emptyset = \emptyset$.
(k) $A \subset B$ if and only if $A \cup B = B$.
(l) $A \subset B$ if and only if $A \cap B = A$.
(m) $A \cup B = B$ if and only if $A \cap B = A$.
(n) $A \subset C(B)$ if and only if $A \cap B = \emptyset$.
(o) $C(A) \subset B$ if and only if $A \cup B = S$.
(p) $A \subset B$ if and only if $C(B) \subset C(A)$.
(q) $A \subset C(B)$ if and only if $B \subset C(A)$.

2. Let $X \subset Y \subset Z$. Prove the following:
   (a) $C_Y(X) \subset C_Z(X)$.
   (b) $Z - (Y - X) = X \cup (Z - Y)$.

4 Indexed Families of Sets

Let I be a set. For each $\alpha \in I$, let A_α be a subset of a given set S. We call I an indexing set and the collection of subsets of S indexed by the elements of I is called an *indexed family* of subsets of S. We denote this indexed family of subsets of S by $(A_\alpha)_{\alpha \in I}$ (read, "A sub-alpha, alpha in I").

The primary purpose for introducing the concept of indexed family of subsets is to allow for a more general formation of unions and intersections of sets. Let $(A_\alpha)_{\alpha \in I}$ be an indexed family of subsets of a set S. The union of this indexed family, written,

$$\bigcup_{\alpha \in I} A_\alpha,$$

(read "union over α in I of A_α") is the set of all elements $x \in S$ such that $x \in A_\beta$ for at least one index $\beta \in I$. The intersection of this indexed family, written,

$$\bigcap_{\alpha \in I} A_\alpha,$$

(read "intersection over α in I of A_α") is the set of all elements $x \in S$ such that $x \in A_\beta$ for each $\beta \in I$. [Note that

$$\bigcup_{\alpha \in I} A_\alpha = \bigcup_{\gamma \in I} A_\gamma,$$

for which reason the two occurrences of "α" in the expression $\bigcup_{\alpha \in I} A_\alpha$ are referred to as dummy indices.]

As an example, let A_1, A_2, A_3, A_4 be respectively the set of freshmen, sophomores, juniors, and seniors in some specified

college. Here we have $I = \{1, 2, 3, 4\}$ as an indexing set and $\bigcup_{\alpha \in I} A_\alpha$ is the set of undergraduates whereas $\bigcap_{\alpha \in I} A_\alpha = \emptyset$.

If the indexing set I contains precisely two distinct indices, then the union over α in I of A_α is the same as the union of two sets as defined in the previous section; that is,

$$\bigcup_{\alpha \in \{i,j\}} A_\alpha = A_i \cup A_j.$$

Similarly,

$$\bigcap_{\alpha \in \{i,j\}} A_\alpha = A_i \cap A_j.$$

DeMorgan's laws are applicable to unions and intersections of indexed families of subsets of a set S.

Theorem Let $(A_\alpha)_{\alpha \in I}$ be an indexed family of subsets of a set S. Then

$$C\Big(\bigcup_{\alpha \in I} A_\alpha\Big) = \bigcap_{\alpha \in I} C(A_\alpha), \tag{4.1}$$

$$C\Big(\bigcap_{\alpha \in I} A_\alpha\Big) = \bigcup_{\alpha \in I} C(A_\alpha). \tag{4.2}$$

Proof. Suppose $x \in C(\bigcup_{\alpha \in I} A_\alpha)$. Then $x \notin \bigcup_{\alpha \in I} A_\alpha$; that is, $x \notin A_\beta$ for each index $\beta \in I$. Thus $x \in C(A_\beta)$ for each index $\beta \in I$ and $x \in \bigcap_{\alpha \in I} C(A_\alpha)$. Therefore,

$$C\Big(\bigcup_{\alpha \in I} A_\alpha\Big) \subset \bigcap_{\alpha \in I} C(A_\alpha).$$

Conversely, suppose that $x \in \bigcap_{\alpha \in I} C(A_\alpha)$. Then $x \in C(A_\beta)$ fo each index $\beta \in I$. Thus $x \notin A_\beta$ for each index $\beta \in I$; that is, $x \notin \bigcup_{\alpha \in I} A_\alpha$. Therefore, $x \in C(\bigcup_{\alpha \in I} A_\alpha)$ and

$$\bigcap_{\alpha \in I} C(A_\alpha) \subset C\Big(\bigcup_{\alpha \in I} A_\alpha\Big).$$

This proves 4.1. The proof of 4.2 is left as an exercise.

Given any collection of subsets of a set S, the concept of indexed family of subsets allows us to define the union or intersection of the aforementioned subsets. We need only construct some convenient indexing set. In the event that the collection of subsets is finite, the finite set $\{1, 2, \ldots, n\}$ of integers is a convenient indexing set. Given subsets $A_1, A_2, \ldots, A_n$ of

S, we shall often write $A_1 \cup A_2 \cup \ldots \cup A_n$ or $\bigcup_{i=1}^{n} A_i$ for $\bigcup_{\alpha \in \{1,2,\ldots,n\}} A_\alpha$ and, similarly, $A_1 \cap A_2 \cap \ldots \cap A_n$ or $\bigcap_{i=1}^{n} A_i$ for $\bigcap_{\alpha \in \{1,2,\ldots,n\}} A_\alpha$.

Exercises

1. Let $(A_\alpha)_{\alpha \in I}$, $(B_\alpha)_{\alpha \in I}$ be two indexed families of subsets of a set S. Prove the following:

(a) For each $\beta \in I$, $A_\beta \subset \bigcup_{\alpha \in I} A_\alpha$.

(b) For each $\beta \in I$, $\bigcap_{\alpha \in I} A_\alpha \subset A_\beta$.

(c) $\bigcup_{\alpha \in I} (A_\alpha \cup B_\alpha) = (\bigcup_{\alpha \in I} A_\alpha) \cup (\bigcup_{\alpha \in I} B_\alpha)$.

(d) $\bigcap_{\alpha \in I} (A_\alpha \cap B_\alpha) = (\bigcap_{\alpha \in I} A_\alpha) \cap (\bigcap_{\alpha \in I} B_\alpha)$.

(e) If for each $\beta \in I$, $A_\beta \subset B_\beta$ then

$$\bigcup_{\alpha \in I} A_\alpha \subset \bigcup_{\alpha \in I} B_\alpha,$$
$$\bigcap_{\alpha \in I} A_\alpha \subset \bigcap_{\alpha \in I} B_\alpha.$$

(f) Let $D \subset S$. Then

$$\bigcup_{\alpha \in I} (A_\alpha \cap D) = (\bigcup_{\alpha \in I} A_\alpha) \cap D,$$
$$\bigcap_{\alpha \in I} (A_\alpha \cup D) = (\bigcap_{\alpha \in I} A_\alpha) \cup D.$$

2. Let $A, B, D \subset S$. Then

$$A \cap (B \cup D) = (A \cap B) \cup (A \cap D),$$
$$A \cup (B \cap D) = (A \cup B) \cap (A \cup D).$$

3. Let $(A_\alpha)_{\alpha \in I}$ be an indexed family of subsets of a set S. Let $J \subset I$. Prove that

(a) $\bigcap_{\alpha \in J} A_\alpha \supset \bigcap_{\alpha \in I} A_\alpha$.

(b) $\bigcup_{\alpha \in J} A_\alpha \subset \bigcup_{\alpha \in I} A_\alpha$.

4. Let $(A_\alpha)_{\alpha \in I}$ be an indexed family of subsets of a set S. Let $B \subset S$. Prove that

(a) $B \subset \bigcap_{\alpha \in I} A_\alpha$ if and only if for each $\beta \in I$, $B \subset A_\beta$.

(b) $\bigcup_{\alpha \in I} A_\alpha \subset B$ if and only if for each $\beta \in I$, $A_\beta \subset B$.

6. Let I be the set of real numbers that are greater than 0. For each $x \in I$, let A_x be the open interval $(0, x)$. Prove that $\bigcap_{x \in I} A_x = \emptyset$, $\bigcup_{x \in I} A_x = I$. For each $x \in I$, let B_x be the closed interval $[0, x]$. Prove that $\bigcap_{x \in I} B_x = \{0\}$, $\bigcup_{x \in I} B_x = I \cup \{0\}$.

5 Products of Sets

Let x and y be objects. The *ordered pair* (x, y)* is a sequence of two objects, the first object of the sequence being x and the second object of the sequence being y. Let A and B be sets. The *Cartesian product* of A and B, written

$$A \times B,$$

(read "A cross B") is the set whose elements are all the ordered pairs (x, y) such that $x \in A$ and $y \in B$.

The Cartesian product of two sets is a familiar notion. The coordinate plane of analytical geometry is the Cartesian product of two lines. The possible outcomes of the throw of a pair of dice is the Cartesian product of two sets, each of which is comprised of the numbers 1, 2, 3, 4, 5, 6. The possible seat designations in many theatres is the Cartesian product of the alphabet and a finite set of integers. Unless $A = B$, the two Cartesian products $A \times B$ and $B \times A$ are distinct.

A generalization of the Cartesian product of two sets is the direct product of a sequence of sets. Let $A_1, A_2, \ldots, A_n$ be a finite sequence of sets, indexed by $\{1, 2, \ldots, n\}$. The *direct product* of $A_1, A_2, \ldots, A_n$, written

$$\prod_{i=1}^{n} A_i$$

* If x and y are real numbers, the symbol (x, y) is ambiguous, for it may stand for either the ordered pair whose first element is x and the second y, or for the open interval (x, y). It is hoped that this ambiguity will be resolved by the context in which the symbol occurs.

(read, "product i equals one to n of A_i") is the set consisting of all sequences $(a_1, a_2, \ldots, a_n)$ such that $a_1 \in A_1$, $a_2 \in A_2$, $\ldots$, $a_n \in A_n$. In particular,

$$\prod_{i=1}^{2} A_i = A_1 \times A_2.$$

For this reason we shall often write

$$A_1 \times A_2 \times \ldots \times A_n$$

for $\prod_{i=1}^{n} A_i$.

The concept of direct product may be extended to an infinite sequence $A_1, A_2, \ldots, A_n, \ldots$ of sets, indexed by the positive integers. The *direct product* of $A_1, A_2, \ldots, A_n, \ldots$, written

$$\prod_{i=1}^{\infty} A_i$$

or

$$A_1 \times A_2 \times \ldots \times A_n \times \ldots$$

is the set whose elements are all infinite sequences $(a_1, a_2, \ldots, a_n, \ldots)$ such that $a_i \in A_i$ for each positive integer i.

The set of points of Euclidean n-space yields an example of a direct product of sets. If for $i = 1, 2, \ldots, n$ we have $A_i = R$, where R is the set of real numbers, then

$$R^n = \prod_{i=1}^{n} A_i$$

is the set of points of a Euclidean n-space. An element $x \in R^n$ is a sequence $x = (x_1, x_2, \ldots, x_n)$ of real numbers. In general, if the sets $A_1, A_2, \ldots, A_n$ are all equal to the same set A, we write

$$A^n = \prod_{i=1}^{n} A_i$$

and call an element $a = (a_1, a_2, \ldots, a_n) \in A^n$ an n-tuple.

Exercises

1. Let $X \subset A$, $Y \subset B$. Prove that $C(X \times Y) = (C(X) \times Y) \cup (X \times C(Y)) \cup (C(X) \times C(Y))$.
2. For each set A, $A \times \emptyset = \emptyset \times A = \emptyset$.
3. Let $X, X' \subset A$; $Y, Y' \subset B$. Prove that if $X \subset X'$ and $Y \subset Y'$ then $X \times Y \subset X' \times Y'$. Show by constructing an example that if $Y = Y' = \emptyset$, then $X \times Y \subset X' \times Y'$ does not imply that $X \subset X'$. Prove that if $X \neq \emptyset$ and $Y \neq \emptyset$ then $X \times Y \subset X' \times Y'$ does imply that $X \subset X'$ and $Y \subset Y'$.
4. Prove that if A has precisely n distinct elements and B has precisely m distinct elements, where m and n are positive integers, then $A \times B$ has precisely mn distinct elements.
5. Let A and B be sets, both of which have at least two distinct members. Prove that there is a subset $W \subset A \times B$ that is not the Cartesian product of a subset of A with a subset of B. [Thus, not every subset of a Cartesian product is the Cartesian product of a pair of subsets.]

6 Functions

The most familiar example of a function in mathematics is a correspondence that associates with each real number x a real number $f(x)$. The basic idea that is to be expressed by the use of the term *function* is the idea of correspondence. For example, the purpose of marking an examination may be described as the construction of a marking function that makes correspond to each student taking the examination some integer between zero and one hundred. Integration of a continuous function defined on some closed interval $[a, b]$ is another example of a function, namely the correspondence that associates with each object f in this given set of objects the real number

$$\int_a^b f(x)\, dx.$$

The concept of function or correspondence need not be restricted to the realm of numerical quantities. The correspondence that associates with each undergraduate in college one of the four adjectives *freshman, sophomore, junior*, or *senior* is also an example of a function.

Definition Let A and B be sets. A correspondence that associates with each element $x \in A$ an element $f(x) \in B$ is called a *function* from A to B, which we shall write

$$f: A \to B,$$

or

$$A \xrightarrow{f} B.$$

Definition Let $f: A \to B$. The subset $\Gamma_f \subset A \times B$, which consists of all ordered pairs of the form $(a, f(a))$ is called the *graph* of $f: A \to B$.

Let A and B be sets. Given a subset Γ of $A \times B$ there is a function $f: A \to B$ such that Γ is the graph of $f: A \to B$ if, for each $x \in A$, there is one and only one element of the form $(x, y) \in \Gamma$.

Definition Let $f: A \to B$ be given. For each subset X of A, the subset of B whose elements are the points $f(x)$ such that $x \in X$ is denoted by $f(X)$. $f(X)$ is called the *image* of X. For each subset Y of B, the subset of A whose elements are the points $x \in A$ such that $f(x) \in Y$ is denoted by $f^{-1}(Y)$. $f^{-1}(Y)$ is called the *inverse image* of Y or f *inverse* of Y.

Definition Let $f: A \to B$ be given. A is called the *domain* of f. $f(A)$ is called the *range* of f.

Example: Let $f: R \to R$, R the set of real numbers, be the function such that for each $x \in R$,

$$f(x) = x^2 - x - 2.$$

If X is the closed interval $[1, 2]$, then $f(X) = [-2, 0]$. If Z is

the open interval $(-1, 1)$, then $f(Z) = (-9/4, 0) \cup \{-9/4\}$. $f^{-1}([-2, 0]) = [1, 2] \cup [-1, 0]$. $f^{-1}(\{0\}) = \{2, -1\}$ is the set of roots of the polynomial $x^2 - x - 2$. $f^{-1}([-5, -4]) = \emptyset$.

A function $f: A \to B$ is also called a *mapping* or *transformation* of A into B. We may think of such a function as carrying each point $x \in A$ into its corresponding point $f(x) \in B$, as indicated in Figure 2. Similarly, $f: A \to B$ carries each subset

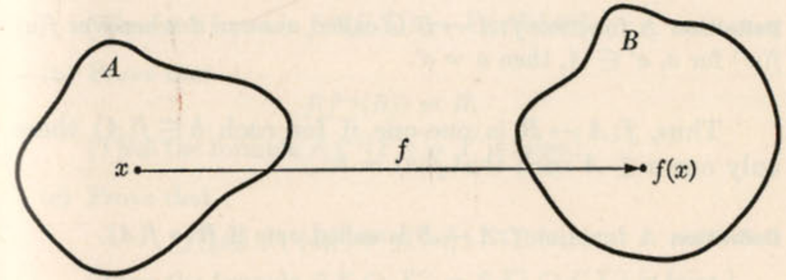


Figure 2

X of A onto the subset $f(X)$ of B, as indicated in Figure 3, and f^{-1} for each subset Y of B is the set of all $x \in A$ that are carried into points of Y, as indicated in Figure 4.

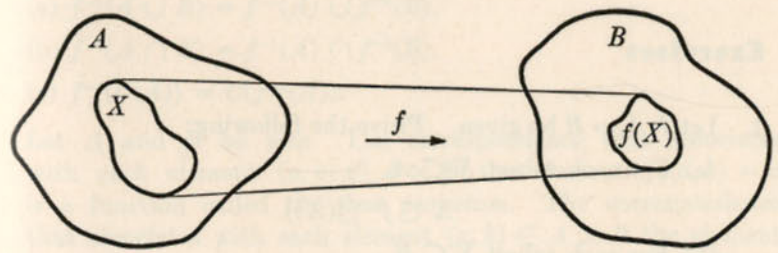


Figure 3

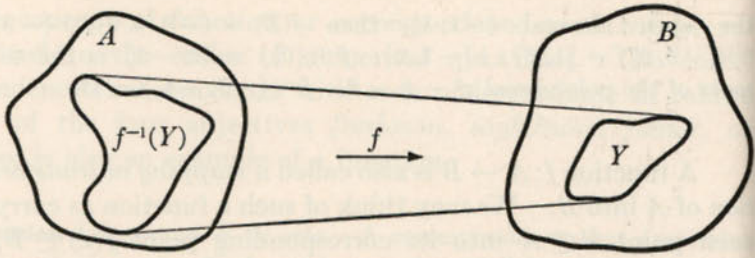


Figure 4

Definition A function $f: A \to B$ is called *one-one* if whenever $f(a) = f(a')$ for $a, a' \in A$, then $a = a'$.

Thus, $f: A \to B$ is one-one if for each $b \in f(A)$ there is only one $a \in A$ such that $f(a) = b$.

Definition A function $f: A \to B$ is called *onto* if $B = f(A)$.

Certain particular types of functions are frequently considered.

Definition A function $f: A \to B$ is called a *constant* function if there is a point $b \in B$ such that $f(x) = b$ for each $x \in A$.

Definition A function $f: A \to A$ is called the *identity* function (on A) if $f(x) = x$ for each $x \in A$.

Exercises

1. Let $f: A \to B$ be given. Prove the following:

(a) For each subset $X \subset A$,
$$X \subset f^{-1}(f(X)).$$

(b) For each subset $Y \subset B$,
$$Y \supset f(f^{-1}(Y)).$$

(c) If $f: A \to B$ is one-one, then for each subset $X \subset A$,
$$f^{-1}(f(X)) = X.$$

(d) If $f: A \to B$ is onto, then for each subset $Y \subset B$,
$$f(f^{-1}(Y)) = Y.$$

2. Let $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$ be two sets, each having precisely two distinct elements. Let $f: A \to B$ be the constant function such that $f(a) = b_1$ for each $a \in A$.

(a) Prove that
$$f^{-1}(f(\{a_1\})) \neq \{a_1\}.$$
[Thus the formula $f^{-1}(f(X)) = X$ is false.]

(b) Prove that
$$f(f^{-1}(B)) \neq B.$$
[Thus the formula $f(f^{-1}(Y)) = Y$ is false.]

(c) Prove that
$$f(\{a_1\} \cap \{a_2\}) \neq f(\{a_1\}) \cap f(\{a_2\}).$$
[Thus the formula $f(X \cap X') = f(X) \cap f(X')$ is false.]

3. Let $f: A \to B$ be given and let X and Y be subsets of A. Prove the following:

(a) $f(X \cup Y) = f(X) \cup f(Y)$.

(b) $f(X \cap Y) \subset f(X) \cap f(Y)$.

(c) If $f: A \to B$ is one-one, then $f(X \cap Y) = f(X) \cap f(Y)$.

4. Given $f: X \to Y$, let A and B be subsets of Y. Prove:

(a) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.

(b) $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

(c) $f^{-1}(C(A)) = C(f^{-1}(A))$.

5. Let A and B be sets. The correspondence that associates with each element $(a, b) \in A \times B$ the element $p_1(a, b) = a$ is a function called the *first projection*. The correspondence that associates with each element $(a, b) \in A \times B$ the element $p_2(a, b) = b$ is a function called the *second projection*. Prove that if $B \neq \emptyset$, then $p_1: A \times B \to A$ is onto and if $A \neq \emptyset$, then

$p_2: A \times B \to B$ is onto. Under what circumstances is p_1 or p_2 one-one? What is $p_1^{-1}(\{a\})$ for an element $a \in A$?

6. Let A and B be sets, with $B \neq \emptyset$. For each $b \in B$ the correspondence that associates with each element $a \in A$ the element $j_b(a) = (a, b) \in A \times B$ is a function. Prove that for each $b \in B$, $j_b: A \to A \times B$ is one-one. Under what circumstances is j_b onto? What is $j_b^{-1}(W)$ for a subset $W \subset A \times B$?

7 Composition of Functions and Diagrams

Definition Let $f: A \to B$ and $g: B \to C$ be given. The *composition* of $f: A \to B$ and $g: B \to C$ is the correspondence that associates with each element $a \in A$, the element $g(f(a)) \in C$. This function is written

$$gf: A \to C,$$

or

$$A \xrightarrow{gf} C.$$

A function $h: A \to C$ is, therefore, the composition of $f: A \to B$ and $g: B \to C$ (often abbreviated by writing $h = gf$) if for each $a \in A$, $h(a) = g(f(a))$. In a pictorial representation of these functions, we have $h = gf$ when these functions behave in the manner indicated in Figure 5.

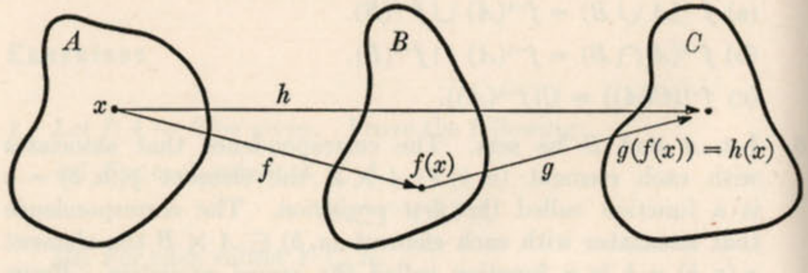


Figure 5

The concept of the composition of functions is extended to the composition of a finite number of functions.

Definition Let $f_1: A_1 \to A_2, f_2: A_2 \to A_3, \ldots, f_n: A_n \to A_{n+1}$ be given. The *composition* of $f_1: A_1 \to A_2, f_2: A_2 \to A_3, \ldots$, and $f_n: A_n \to A_{n+1}$ is the correspondence that associates with each element $x \in A_1$ the element $f_n(\ldots f_2(f_1(x)) \ldots) \in A_{n+1}$. This function is written

$$f_n \cdots f_2 f_1: A_1 \to A_{n+1},$$

or

$$A_1 \xrightarrow{f_n \cdots f_2 f_1} A_{n+1}.$$

Let three functions $f: A \to B$, $g: B \to C$, and $h: C \to D$ be given. We may form $hgf: A \to D$. We may also form $gf: A \to C$ and compose this function with $h: C \to D$ to obtain $h(gf): A \to D$. Similarly, we may form $(hg)f: A \to D$. We thus have three functions hgf, $h(gf)$, $(hg)f: A \to D$. But

$$(hgf)(x) = h(g(f(x)));$$
$$(h(gf))(x) = h((gf)(x)) = h(g(f(x)));$$
$$((hg)f)(x) = (hg)(f(x)) = h(g(f(x))).$$

Thus, these three functions are the same. This observation provides a basis for the justification of the removal or replacement of parentheses in expressions such as $(f_4 f_3)(f_2 f_1)$, etc.

Suppose we are given three functions $f: A \to B$, $g: B \to C$, and $k: A \to C$. The existence of these three functions may be indicated, as in Figure 6, by what we shall call a *diagram*. The letters A, B, C stand for the various sets, and an arrow leading from one set to another indicates a function from the first set to the second, namely, the function that carries each

$$\begin{tikzcd} A \arrow[rr, "k"] \arrow[dr, "f"'] & & C \\ & B \arrow[ur, "g"'] & \end{tikzcd}$$

Figure 6

element x of the first set into the element $t(x)$ of the second set, where t stands for the symbol closest to the middle of the arrow. The fact that we may form the composition of two functions (such as $gf:A \to C$ in the above diagram) is represented by a path in the direction of the arrows that goes from one set to a second and from the second set to a third. (In the above diagram we say, "We may go from A to B via f and from B to C via g.")

We shall desire to diagram more complex situations than the one indicated in Figure 6. Let us say that by a *diagram* we shall mean a figure consisting of several symbols denoting sets and arrows leading from one symbol to another, each arrow leading from a set X to a set Y having an associated symbol t, the arrow and its symbol representing a given function $t:X \to Y$. For example, diagram (7.1) indicates the existence of given functions $f:A \to B$, $g:A \to C$,

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "k"] \\ C \arrow[r, "h"] & D \end{tikzcd} \tag{7.1}$$

$k:B \to D$, $h:C \to D$. This diagram shows us that by composing functions we may obtain two functions from A to D, namely, kf, $hg:A \to D$. In any diagram, a path from X to Y consisting of a sequence of arrows leading from X to Y indicates the existence of a function from X to Y obtained by composing the functions represented by these arrows in the order of their occurrence, starting at X and terminating at Y.

In diagram (7.1) it may or may not be true that $kf = hg$. In the event that $kf = hg$ we will say that diagram (7.1) is *commutative*. In general, a diagram is said to be *commutative* if for each X and Y in the diagram that represent sets, and for any two paths in the diagram beginning at X and ending at Y,

the two functions from X to Y so represented are equal. For example, the statement that diagram (7.2) is commutative means that $f = jh$, $k = gj$, and $kh = gjh = gf$, (note that the first two equalities imply the third).

$$\begin{tikzcd} & D \arrow[d, "j"] \arrow[dr, "k"] & \\ A \arrow[ur, "h"] \arrow[r, "f"] & B \arrow[r, "g"] & C \end{tikzcd} \tag{7.2}$$

A given set A may occur more than once in a diagram. For example, let A be the set of positive real numbers and R the set of real numbers. Let $f:A \to R$ be defined by the correspondence $f(x) = \log_e x$, $x \in A$, and let $g:R \to A$ be defined by the correspondence $g(x) = e^x$, $x \in R$. Let $i:A \to A$ be the identity function. Then the diagram (7.3) is commutative, for $(gf)(x) = e^{\log_e x} = x = i(x)$ for $x \in A$.

$$\begin{tikzcd} A \arrow[rr, "i"] \arrow[dr, "f"'] & & A \\ & R \arrow[ur, "g"'] & \end{tikzcd} \tag{7.3}$$

Exercises

1. Using the functions defined by the correspondences $g(x) = x^2$ and $h(x) = \sqrt{x}$, $x \geqq 0$, construct an example of a commutative diagram similar to diagram (7.3).

2. Let $f:R \times R \to R$ be the function defined by the correspondence $f(x, y) = x^2 + y^2$ and let $g:R \times R \to R$ be the function defined by the correspondence $g(x, y) = x + y$. Let $h:R \to R$ be the func-

tion defined by the correspondence $h(x) = x^2$. Is the diagram

$$\begin{tikzcd} R \times R \arrow[rr, "f"] \arrow[dr, "g"'] & & R \\ & R \arrow[ur, "h"'] & \end{tikzcd}$$

commutative?

3. Let S be a non-empty set. Let $f: 2^S \times 2^S \to 2^S$ be the function defined by the correspondence $f(A, B) = A \cup B$ for $A, B \subset S$. Let $g: 2^S \times 2^S \to 2^S$ be the function defined by the correspondence $g(A, B) = A \cap B$ for $A, B \subset S$. Let $h: 2^S \to 2^S$ be the function defined by the correspondence $h(A) = C(A)$ for $A \subset S$. Is the diagram

$$\begin{tikzcd} 2^S \times 2^S \arrow[rr, "f"] \arrow[dr, "g"'] & & 2^S \\ & 2^S \arrow[ur, "h"'] & \end{tikzcd}$$

commutative?

Let $k: 2^S \times 2^S \to 2^S \times 2^S$ be the function defined by the correspondence $k(A, B) = (C(A), C(B))$ for $A, B \subset S$. Is the diagram

$$\begin{tikzcd} 2^S \times 2^S \arrow[r, "f"] \arrow[d, "k"] & 2^S \arrow[d, "h"] \\ 2^S \times 2^S \arrow[r, "g"] & 2^S \end{tikzcd}$$

commutative?

4. Let $f: R \to R$ be the function defined by the correspondence $f(x) = x^2$ and let $g: R \to R$ be the function defined by the correspondence $g(x) = x + 1$. Is the diagram

$$\begin{tikzcd} R \arrow[r, "f"] \arrow[d, "g"] & R \arrow[d, "g"] \\ R \arrow[r, "f"] & R \end{tikzcd}$$

commutative?

5. Let $i: A \to A$ be the identity function. Let the diagram

$$\begin{tikzcd} A \arrow[rr, "i"] \arrow[dr, "f"'] & & A \\ & B \arrow[ur, "g"'] & \end{tikzcd}$$

be commutative. Prove that $g: B \to A$ is onto and that $f: A \to B$ is one-one.

6. Let $f: A \to B$ be one-one and let $i: A \to A$ be the identity function. Define a function $g: B \to A$ such that the diagram

$$\begin{tikzcd} A \arrow[rr, "i"] \arrow[dr, "f"'] & & A \\ & B \arrow[ur, "g"'] & \end{tikzcd}$$

is commutative.

7. Let $g: B \to A$ be onto and let $i: A \to A$ be the identity function. Define a function $f: A \to B$ such that the diagram

$$\begin{tikzcd} A \arrow[rr, "i"] \arrow[dr, "f"'] & & A \\ & B \arrow[ur, "g"'] & \end{tikzcd}$$

is commutative.

8. Let $f:A \to B$, $g:B \to C$. Prove that for $Z \subset C$, $(gf)^{-1}(Z) = f^{-1}(g^{-1}(Z))$.

8 Inverse Functions

Definition Let $f:A \to B$ and $g:B \to A$ be given. The function $f:A \to B$ is called the *inverse* of $g:B \to A$ and the function $g:B \to A$ is called the *inverse* of $f:A \to B$ if

$$g(f(a)) = a$$

for each $a \in A$ and

$$f(g(b)) = b$$

for each $b \in B$.

In this event we shall also say that $f:A \to B$ and $g:B \to A$ are *inverse functions* and that each of them is *invertible.*

Let $i_A:A \to A$ and $i_B:B \to B$ be identity functions. The statement that $f:A \to B$ and $g:B \to A$ are inverse functions is equivalent to the statement that the two diagrams

$$\begin{tikzcd} A \arrow[rr, "i_A"] \arrow[dr, "f"'] & & A \\ & B \arrow[ur, "g"'] & \end{tikzcd} \qquad \begin{tikzcd} B \arrow[rr, "i_B"] \arrow[dr, "g"'] & & B \\ & A \arrow[ur, "f"'] & \end{tikzcd}$$

are commutative.

Theorem Let $f:A \to B$ and $g:B \to A$ be inverse functions, then both functions are one-one and onto.

Proof. Suppose $f(x) = f(y)$, x, $y \in A$. Then $x = g(f(x)) = g(f(y)) = y$ and therefore f is one-one. To show that f is onto, let $b \in B$. We have $g(b) \in A$ and $f(g(b)) = b$, therefore if we set $a = g(b)$, $b = f(a)$ and f is onto. The roles of the two functions may be interchanged, since the definition of inverse functions imposes conditions symmetrical with regard to the two functions. Therefore, $g:B \to A$ is also one-one and onto.

We have shown that, given a function $h:X \to Y$, a necessary condition that this function be invertible is that the function be one-one and onto. This condition is also sufficient.

Theorem Let $f:A \to B$ be one-one and onto. Then there exists a function $g:B \to A$ such that these two functions are inverse functions.

Proof. We shall first define $g:B \to A$. Given $b \in B$, we may write $b = f(a)$ for some $a \in A$ since f is onto. Furthermore, f is one-one; hence there is only one element $a \in A$ such that $f(a) = b$. We define $g(b) = a$. The correspondence that associates with each $b \in B$ the element $g(b) \in A$, as defined above, is a function $g:B \to A$. $f(g(b)) = b$ for each $b \in B$ by the definition of $g:B \to A$. Given $a \in A$, let $a' = g(f(a))$. Then $f(a') = f(g(f(a))) = f(a)$ by the remark just made. Since $f:A \to B$ is one-one, $a = a' = g(f(a))$. Thus, $f:A \to B$ and $g:B \to A$ are inverse functions.

The last two theorems may be combined in the statement: given $f:A \to B$, a necessary and sufficient condition that there be a function $g:B \to A$ such that these two functions are inverse functions is that $f:A \to B$ be one-one and onto. Furthermore, in this event, the function $g:B \to A$ is uniquely determined.

Theorem Let $f:A \to B$, $g:B \to A$ be inverse functions and let $f:A \to B$ and $g':B \to A$ be inverse functions. Then $g:B \to A$ and $g':B \to A$ are equal.

Proof. We must prove that $g(b) = g'(b)$ for each $b \in B$. But $b = f(g(b))$ and therefore $g'(b) = g'(f(g(b))) = g(b)$, since $g'(f(a)) = a$ for each $a \in A$.

The proof of this last theorem may also be viewed as a direct consequence of the commutativity of the diagram

$$\begin{tikzcd} B \arrow[rr, "i_B"] \arrow[dr, "g"'] & & B \arrow[dr, "g'"] \\ & A \arrow[ur, "f"] \arrow[rr, "i_A"] & & A \end{tikzcd}$$

which yields $g'(b) = g'(i_B(b)) = g'(f(g(b))) = i_A(g(b)) = g(b)$.

Exercises

1. Let A be the set of positive real numbers. For each $a \in A$, let $f_a : R \to A$ be the function defined by the correspondence $f_a(x) = a^x$, $x \in R$, and let $g_a : A \to R$ be the function defined by the correspondence $g_a(x) = \log_a x$, $x \in A$. Prove that for each $a \in A$, $f_a : R \to A$ and $g_a : A \to R$ are inverse functions.
2. Let $f : [-1, 1] \to R$ be the function defined by the correspondence $f(x) = \arcsin x$, $x \in [-1, 1]$ and let $g : R \to [-1, 1]$ be the function defined by the correspondence $g(x) = \sin x$, $x \in R$. Prove that these two functions are not inverse functions.
3. Let $f : [-1, 1] \to [-\pi/2, \pi/2]$ be the function defined by the correspondence $f(x) = \arcsin x$, $x \in [-1, 1]$ and let $g : [-\pi/2, \pi/2] \to [-1, 1]$ be the function defined by the correspondence $g(x) = \sin x$, $x \in [-\pi/2, \pi/2]$. Prove that these two functions are inverse functions.
4. Let A be the set of all functions $f : [a, b] \to R$ that are continuous on $[a, b]$. Let B be the subset of A consisting of all functions possessing a continuous derivative on $[a, b]$. Let C be the subset of B consisting of all functions whose value at a is 0. Let $d : B \to A$ be the correspondence that associates with each function in B its derivative. Is the function $d : B \to A$ invertible?

   To each $f \in A$, let $h(f)$ be the function defined by

   $$(h(f))(x) = \int_a^x f(t)\,dt,$$

   for $x \in [a, b]$. Verify that $h : A \to C$. Find the function $g : C \to A$ such that these two functions are inverse functions.

9 Restriction and Extension of Functions

Definition Let $A \subset X$. Let $f : A \to Y$ and $F : X \to Y$. If for each $x \in A$, $f(x) = F(x)$, we say that *F is an extension of f to X* or that *f is a restriction of F to A*. In this event we shall write

$$f = F \mid A.$$

Examples Let A be the open interval $(0, \pi/2)$. For each $\theta \in A$, let Δ_θ be a right triangle one of whose acute angles is θ radians, and let $f(\theta)$ be the ratio of the length of the side of this triangle opposite the angle of magnitude θ to the length of the hypotenuse of Δ_θ, (more familiarly,

$$f(\theta) = \frac{\text{opposite}}{\text{hypotenuse}}\Big)$$

Thus $f : A \to R$. For each $\theta \in R$, let $(a, b)_\theta$ be the point of the plane R^2 whose distance from the origin is 1 and such that the rotation about the origin of the line segment whose end points are the origin and $(1, 0)$ to the position of the line segment whose end points are the origin and $(a, b)_\theta$ represents an angle of magnitude θ radians. Define $F(\theta) = b$. Then $F : R \to R$. F is an extension of f to R as is easily seen if one recognizes $f : A \to R$ as the sine function defined for acute angles by means of right triangles and $F : R \to R$ as the sine function defined for angles of arbitrary magnitude by means of the unit circle.

As a second, less familiar example, let A be the set of positive integers and X the set of real numbers greater than -1. Define $f : A \to R$ by the correspondence

$$f(n) = n!,\ n \in A.$$

Define $F : X \to R$ by the correspondence

$$F(x) = \int_0^\infty t^x e^{-t}\,dt,\ x \in X.$$

[$F(x)$ is related to the Gamma function; in fact, $F(x) = \Gamma(x + 1)$.]

$$F(1) = \int_0^\infty te^{-t}\,dt = 1 = f(1).$$

Integration by parts shows that for $x > 0$, $F(x) = xF(x - 1)$ and provides the basis for an inductive argument that establishes the formula $F(n) = n! = f(n)$ for $n \in A$. Therefore, F is an extension of f to X.

Definition Let $A \subset X$. The function $i:A \to X$, which is defined by the correspondence $i(x) = x$ for each $x \in A$ is called an *inclusion* mapping or function.

Let $A \subset X$, $f:A \to Y$ and $F:X \to Y$. Then F is an extension of f if and only if the diagram

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "i"'] & & Y \\ & X \arrow[ur, "F"'] & \end{tikzcd}$$

is commutative, where $i:A \to X$ is an inclusion mapping.

Given $F:X \to Y$, there are as many restrictions of $F:X \to Y$ as there are subsets of X. Given a subset $A \subset X$, we may obtain the restriction of F to A by forming the composition of the inclusion mapping $i:A \to X$ and $F:X \to Y$. Thus, we may write

$$F \mid A = Fi.$$

Exercises

1. Let $A \subset B \subset X$. Let $i:A \to B$ and $j:B \to X$ be inclusion mappings. Prove that $ji:A \to X$ is an inclusion mapping.
2. Let $A \subset B \subset X$. Let $f:A \to Y$, $g:B \to Y$, and $F:X \to Y$. Prove that if g is an extension of f to B and F is an extension of g to X, then F is an extension of f to X.

3. Let m, n be positive integers. Let X be a set with m distinct elements and Y a set with n distinct elements. How many distinct functions are there from X to Y? Let A be a subset of X with r distinct elements, $0 \leqq r < m$ and $f:A \to Y$. How many distinct extensions of f to X are there?
4. Describe an exponential function in terms of extensions of a function $f:A \to R$, where A is the set of positive integers.

Metric Spaces

II

1 Introduction

A metric space is a set of points and a prescribed quantitative measure of the degree of closeness of pairs of points in this space. The real number system and the coordinate plane of analytical geometry are familiar examples of metric spaces. Starting from the vague characterization of a continuous function as one that transforms nearby points into points that are themselves nearby, we can, in a metric space, formulate a precise definition of continuity. Although this definition may be stated in the so-called "ε, δ" terminology, there are other, equivalent formulations available in a metric space. These include characterizations of continuity in terms of the behavior of a function with respect to certain subsets called neighborhoods of a point, or with respect to certain subsets called open sets.

2 Metric Spaces

Given two real numbers a and b, there is determined a non-negative real number, $|a - b|$, called the distance between a and b. Since to each ordered pair (a, b) of real numbers there is associated the real number $|a - b|$, we may write this correspondence in functional notation by setting

$$d(a, b) = |a - b|.$$

Thus we have a function $d: R \times R \to R$, where R is the set of real numbers. This function has four important properties, which the reader should verify:

1. $d(x, y) \geqq 0$;
2. $d(x, y) = 0$ if and only if $x = y$;
3. $d(x, y) = d(y, x)$;
4. $d(x, z) \leqq d(x, y) + d(y, z)$;

for $x, y, z \in R$. For the purposes of discussing "continuity" of functions, these four properties of "distance" are sufficient. This fact suggests the possibility of examining "continuity" in a more general setting; namely, in terms of any set of points for which there is defined a "distance function" such as the function $d: R \times R \to R$ above.

Definition 2.1 A pair of objects (X, d) consisting of a non-empty set X and a function $d: X \times X \to R$, where R is the set of real numbers, is called a *metric space* provided that:

1. $d(x, y) \geqq 0, \quad x, y \in X$;
2. $d(x, y) = 0$ if and only if $x = y, \quad x, y \in X$;
3. $d(x, y) = d(y, x), \quad x, y \in X$;
4. $d(x, z) \leqq d(x, y) + d(y, z), \quad x, y, z \in X$.

The function d is called a *distance function* on X and the set X is called the *underlying set*.

[A more precise notation for a metric space would be $(X, d: X \times X \to R)$ and for a distance function $d: X \times X \to R$.

We shall, however, frequently delete the sets and arrow in the symbol for a function, when, in a given context, it is clear which sets are involved.]

We may think of the distance function d as providing a quantitative measure of the degree of closeness of two points. In particular, the inequality $d(x, z) \leqq d(x, y) + d(y, z)$ may be thought of as asserting the transitivity of closeness; that is, if x is close to y and y is close to z, then x is close to z.

The verification of the four enumerated properties of the function defined by the correspondence

$$d(a, b) = |a - b|,$$

for $a, b \in R$, R the set of real numbers, therefore proves:

Theorem 2.2 (R, d) is a metric space, where d is the function defined by the correspondence

$$d(a, b) = |a - b|,$$

for $a, b \in R$.

Given a finite collection $(X_1, d_1), (X_2, d_2), \ldots, (X_n, d_n)$ of metric spaces, there is a standard procedure for converting the set

$$X = \prod_{i=1}^{n} X_i$$

into a metric space; that is, for defining a distance function on X.

Theorem 2.3 Let metric spaces $(X_1, d_1), (X_2, d_2), \ldots, (X_n, d_n)$ be given and set

$$X = \prod_{i=1}^{n} X_i.$$

For each pair of points $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n) \in X$, let

$$d: X \times X \to R$$

be the function defined by the correspondence

$$d(x, y) = \underset{1 \leqq i \leqq n}{\text{maximum}} \{d_i(x_i, y_i)\}.$$

Then (X, d) is a metric space.

Proof. With x and y as above, $d_i(x_i, y_i) \geqq 0$ for $1 \leqq i \leqq n$, and therefore $d(x, y) \geqq 0$. If $d(x, y) = 0$, then $d_i(x_i, y_i) = 0$ for $1 \leqq i \leqq n$ and therefore $x_i = y_i$ for each i. Consequently, $x = y$. Conversely, if $x = y$, then $d_i(x_i, y_i) = 0$ for each i, and $d(x, y) = 0$. Since $d_i(x_i, y_i) = d_i(y_i, x_i)$ for $1 \leqq i \leqq n$, $d(x, y) = d(y, x)$. Finally, let $z = (z_1, z_2, \ldots, z_n) \in X$. Let j and k be integers such that $d(x, y) = d_j(x_j, y_j)$ and $d(y, z) = d_k(y_k, z_k)$. Thus, for $1 \leqq i \leqq n$, $d_i(x_i, y_i) \leqq d_j(x_j, y_j)$ and $d_i(y_i, z_i) \leqq d_k(y_k, z_k)$, and

$$\begin{aligned} d_i(x_i, z_i) \leqq d_i(x_i, y_i) + d_i(y_i, z_i) \leqq d_j(x_j, y_j) + d_k(y_k, z_k) \\ = d(x, y) + d(y, z). \end{aligned}$$

Therefore $d(x, z) = \underset{1 \leqq i \leqq n}{\text{maximum}} \{d_i(x_i, z_i)\} \leqq d(x, y) + d(y, z)$.

As an immediate application of this theorem, we have:

Corollary 2.4 (R^n, d) is a metric space, where $d: R^n \times R^n \to R$ is the function defined by the correspondence

$$\begin{aligned} &d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) \\ &\quad = \underset{1 \leqq i \leqq n}{\text{maximum}} \{|x_i - y_i|\}, (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in R^n. \end{aligned}$$

It is interesting to compare the metric space (R^2, d) that we obtain in the above manner with what might be considered a more natural model of the coordinate plane. In (R^2, d) as defined above, the distance from the point $(1, 2)$ to the point $(3, 1)$ is 2, since maximum $\{|1 - 3|, |2 - 1|\} = 2$. The distance function d' used in analytical geometry would yield

$$d'((1, 2), (3, 1)) = \sqrt{(1 - 3)^2 + (2 - 1)^2} = \sqrt{5}.$$

If, for each pair of points $(x_1, x_2), (y_1, y_2) \in R^2$ we define

$$d'((x_1, x_2), (y_1, y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2},$$

then we are constructing a new metric space (R^2, d'), (provided, of course, that d' is a distance function), which must be distinguished from the metric space (R^2, d) where

$$d((x_1, x_2), (y_1, y_2)) = \text{maximum}\{|x_1 - y_1|, |x_2 - y_2|\}.$$

For example, in (R^2, d) the set M of points x such that $d(x, a) \leqq 1$ for a fixed point $a \in R^2$ is a square of width 2 whose center is at a and whose sides are parallel to the coordinate axes, whereas in (R^2, d') the set of points x such that $d'(x, a) \leqq 1$ for a fixed point $a \in R^2$ is a circular disc whose center is a and whose radius is 1 (see Figure 7).

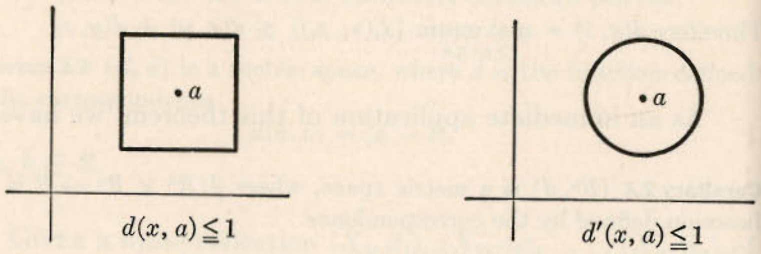


Figure 7

The formula used to define the function d' may be generalized to yield a distance function for R^n.

Theorem 2.5 (R^n, d') is a metric space, where d' is the function defined by the correspondence

$$d'(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

for $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n) \in R^n$.

The proof of this theorem will be found at the end of this chapter.

The fact that we have metric spaces (R^n, d) and (R^n, d'), with d and d' defined as above, serves to emphasize the fact that a metric space consists of two objects, a set and a distance function. Two metric spaces may be distinct even though the underlying sets of points of the two spaces are the same. Given a metric space (X, d), there are some trivial ways of constructing new metric spaces that have the same underlying set of points.

Theorem 2.6 Let (X, d) be a metric space. Let k be a positive real number. Then (X, d_k) is a metric space, where the function d_k is defined by the correspondence

$$d_k(x, y) = k \cdot d(x, y),$$

for $x, y \in X$.

The proof of this theorem is a straightforward verification of the fact that d_k has the four properties of a distance function. The details are left to the reader as an exercise.

Exercises

1. Prove that (R^n, d'') is a metric space, where the function d'' is defined by the correspondence

$$d''(x, y) = \sum_{i=1}^{n} |x_i - y_i|,$$

for $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n) \in R^n$. In (R^2, d'') determine the shape and position of the set of points x such that $d''(x, a) \leqq 1$ for a point $a \in R^2$.

2. Let d and d' be the distance functions defined on R^n in this section and let d'' be the distance function defined in Problem 1 above. Prove that for each pair of points $x, y \in R^n$,

$$d(x, y) \leqq d'(x, y) \leqq \sqrt{n}\, d(x, y),$$
$$d(x, y) \leqq d''(x, y) \leqq n \cdot d(x, y).$$

3. Let X be the set of all continuous functions $f:[a, b] \to R$. For $f, g \in X$, define
$$d(f, g) = \int_a^b |f(t) - g(t)|\, dt.$$
Prove that (X, d) is a metric space.

4. Let $S \subset R$. A function $f:S \to R$ is called *bounded* if there is a real number K such that $|f(x)| \leq K$, $x \in S$ (or equivalently, $f(S) \subset [-K, K]$). Let X' be the set of all bounded functions $f:[a, b] \to R$. For $f, g \in X'$ define
$$d'(f, g) = \text{l.u.b. } \cup_{x \in [a,b]} \{|f(x) - g(x)|\},$$
(l.u.b. is an abbreviation of *least upper bound*). Prove that (X', d') is a metric space.

5. Let $f, g:[a, b] \to R$ be two functions that are both continuous and bounded. Compare $d(f, g)$ and $d'(f, g)$, where d and d' are defined as in Problems 3 and 4 respectively.

6. Let (X, d) be a metric space. For $x, y \in X$ define the function d_B by
$$d_B(x, y) = \frac{d(x, y)}{1 + d(x, y)}.$$
Prove that (X, d_B) is a metric space and that $0 \leq d_B(x, y) < 1$ for $x, y \in X$. [See Hall and Spencer, *Elementary Topology* pp. 88, 89.]

7. Let X be a set. For $x, y \in X$ define the function d by
$$d(x, x) = 0,$$
and
$$d(x, y) = 1,$$
if $x \neq y$. Prove that (X, d) is a metric space.

8. Let Z be the set of integers. Let p be a positive prime integer. Given distinct integers m, n there is a unique integer $t = t(m, n)$ such that
$$m - n = p^t \cdot k,$$
where k is an integer not divisible by p. Define a function $d:Z \times Z \to R$ by the correspondence
$$d(m, m) = 0$$
and
$$d(m, n) = \frac{1}{p^t}$$
from $m \neq n$. Prove that (Z, d) is a metric space. [*Hint:* for $a, b, c \in Z$, $t(a, c) \geq$ minimum $\{t(a, b), t(b, c)\}$]. Let $p = 3$. What is the set of elements $x \in Z$ such that $d(0, x) < 1$? What is the set of elements $x \in Z$ such that $d(0, x) < \frac{1}{3}$?

3 Continuity

In calculus, the first occurrence of the word "continuity" is with reference to a function $f:R \to R$, R the set of real numbers. To decide which condition or conditions this function must satisfy for us to say, "the function f is continuous at a point $a \in R$," we try to decide upon a precise formulation of the statement "a number $f(x)$ will be close to the number $f(a)$ whenever the number x is close to a." Having defined a distance function for the real numbers R, we have a quantitative measure of the degree of closeness of two numbers. But how close must $f(x)$ be to $f(a)$? Instead of specifying some particular degree of closeness of $f(x)$ to $f(a)$, let us think, rather, of requiring that no matter what choice is made for the degree of closeness of $f(x)$ to $f(a)$, it can be so arranged that this degree of closeness is achieved. By the phrase "arrange matters" we mean that we can find a corresponding degree of closeness so that whenever x is within this corresponding degree of closeness to a, then $f(x)$ is within the prescribed degree of closeness to $f(a)$. We have now arrived at the following formulation, "the function $f:R \to R$ is continuous at the number $a \in R$, if given a prescribed degree of closeness, $f(x)$ will be within this prescribed degree of closeness to $f(a)$, whenever x is within some corresponding degree of closeness to a." To put this statement in its final form, we shall substitute for "a prescribed degree of closeness" the symbol "ε," and for the

phrase "some corresponding degree of closeness" the symbol "δ," and use the distance function to measure the degree of closeness.

Definition 3.1 Let $f:R \to R$. The function f is said to be *continuous at the point* $a \in R$, if given $\varepsilon > 0$, there is a $\delta > 0$, such that

$$|f(x) - f(a)| < \varepsilon,$$

whenever

$$|x - a| < \delta.$$

The function f is said to be *continuous* if it is continuous at each point of R.

Because we initially formulated the definition of continuity in terms of the phrase "degree of closeness," we may easily devise a definition of "continuity" applicable to metric spaces in general, since we need only use the distance functions of these metric spaces to measure "degree of closeness."

Definition 3.2 Let (X, d) and (Y, d') be metric spaces, and let $a \in X$. A function $f:X \to Y$ is said to be *continuous at the point* $a \in X$ if given $\varepsilon > 0$, there is a $\delta > 0$, such that

$$d'(f(x), f(a)) < \varepsilon$$

whenever $x \in X$ and

$$d(x, a) < \delta.$$

The function $f:X \to Y$ is said to be *continuous* if it is continuous at each point of X.

Definitions, such as those given above, are created to serve two purposes. First of all, they are abbreviations. Thus, the statement that begins, "given $\varepsilon > 0$, there is . . . ," is replaced by the shorter statement, "$f:X \to Y$ is continuous at the point $a \in X$." Second, these definitions are attempts to formulate precise characterizations of what we feel are significant properties; in this case, the property of being continuous at a point. We have tried to indicate in the discussion preceding these definitions that they do provide a precise characterization of our intuitive, and perhaps vague, concept of continuity. There are, in a certain sense, tests that we may apply to see whether or not they do so. As an illustration, there are certain functions that we "know" are "continuous," that is, we are sure that they possess this property we are trying to characterize. If it should turn out that a function we "know" to be "continuous" is not continuous in accordance with these definitions, then, although these definitions may be precise, they would not furnish a precise characterization of the property we have in mind when we say a function is "continuous." This type of testing of a definition thus takes the form of proving theorems to the effect that certain functions are continuous. For example:

Theorem 3.3 Let (X, d) and (Y, d') be metric spaces. Let $f:X \to Y$ be a constant function, then f is continuous.

Proof. Let a point $a \in X$ and $\varepsilon > 0$ be given. Choose any $\delta > 0$, say $\delta = 1$. Then whenever $d(x, a) < \delta$, we have $d'(f(x), f(a)) = 0 < \varepsilon$.

Theorem 3.4 Let (X, d) be a metric space. Then the identity function $i:X \to X$ is continuous.

Proof. Suppose $a \in X$. Let $\varepsilon > 0$ be given. Choose $\delta = \varepsilon$, then whenever $d(x, a) < \delta$ we have $d(i(x), i(a)) = d(x, a) < \varepsilon$.

Note that in the above proof we could have equally well chosen δ to be any positive number, provided only that $\delta \leqq \varepsilon$ and the proof would still be valid. The choice of δ need not be a very efficient choice; all that is required is that it "do the job." Later theorems will also confirm the fact that our definition of continuity is a "correct" one.

There is one situation we shall have to consider for which the notation $f:X \to Y$ that we have adopted for a function from a metric space (X, d) into a metric space (Y, d') is am-

biguous. Consider metric spaces (X, d) and (X, d') with the same underlying set. If we simply write $f: X \to X$ for a function, it is impossible to tell which metric space is denoted by the first occurrence of X and which by the second. For this reason, when considering one set X with two different distance functions, we shall write

$$f:(X, d) \to (X, d')$$

if we intend to think of $f: X \to X$ as a function from the metric space (X, d) into the metric space (X, d'). As an illustration, we shall prove:

Theorem 3.5 Let $i: R^n \to R^n$ be the identity function. Then

$$i:(R^n, d) \to (R^n, d')$$

and

$$i:(R^n, d') \to (R^n, d)$$

are continuous, where the distance functions d and d' are as defined in Section 2.

Proof. Let $a = (a_1, a_2, \ldots, a_n) \in R^n$. We shall first prove that $i:(R^n, d) \to (R^n, d')$ is continuous. Let $\varepsilon > 0$ be given. Choose $\delta = \varepsilon/\sqrt{n}$. Suppose $x = (x_1, x_2, \ldots, x_n)$ is such that $d(x, a) < \delta$; that is, $\underset{1 \leq i \leq n}{\text{maximum}} \{|a_i - x_i|\} < \delta$. Then

$$d'(x, a) = \sqrt{\sum_{i=1}^{n} (a_i - x_i)^2} < \sqrt{n\delta^2} = \sqrt{\varepsilon^2}.$$

Therefore, given $\varepsilon > 0$, there is a $\delta > 0$ such that $d'(i(x), i(a)) < \varepsilon$ whenever $d(x, a) < \delta$.

We now prove that $i:(R^n, d') \to (R^n, d)$ is continuous. Let $\varepsilon > 0$ be given. Choose $\delta = \varepsilon$. Suppose that $x = (x_1, x_2, \ldots, x_n)$ is such that $d'(x, a) < \delta$. Then

$$\sum_{i=1}^{n} (a_i - x_i)^2 < \delta^2$$

and therefore for each i, $(a_i - x_i)^2 < \delta^2$, or $|a_i - x_i| < \delta = \varepsilon$. Consequently, $d(x, a) < \varepsilon$. Thus, given $\varepsilon > 0$, there is a $\delta > 0$, such that $d(i(x), i(a)) < \varepsilon$ whenever $d'(x, a) < \delta$.

One of the most important elementary theorems about continuous functions is the statement that the composition of two continuous functions is again a continuous function.

Theorem 3.6 Let (X, d), (Y, d'), (Z, d'') be metric spaces. Let $f: X \to Y$ be continuous at the point $a \in X$ and let $g: Y \to Z$ be continuous at the point $f(a) \in Y$. Then $gf: X \to Z$ is continuous at the point $a \in X$.

Proof. Let $\varepsilon > 0$ be given. We must find a $\delta > 0$ such that whenever $x \in X$ and $d(x, a) < \delta$, then $d''(g(f(x)), g(f(a))) < \varepsilon$. Since g is continuous at $f(a)$, there is an $\eta > 0$, such that whenever $y \in Y$ and $d'(y, f(a)) < \eta$, then $d''(g(y), g(f(a))) < \varepsilon$. Using the fact that f is continuous at a, we know that given $\eta > 0$, there is a $\delta > 0$, such that $x \in X$ and $d(x, a) < \delta$ imply that $d'(f(x), f(a)) < \eta$ and hence $d''(g(f(x)), g(f(a))) < \varepsilon$.

Corollary 3.7 Let (X, d), (Y, d'), (Z, d'') be metric spaces. Let $f: X \to Y$ and $g: Y \to Z$ be continuous. Then $gf: X \to Z$ is continuous.

Exercises

1. Let X be the set of continuous functions $f:[a, b] \to R$. Let d^* be the distance function on X defined by

$$d^*(f, g) = \int_a^b |f(t) - g(t)|\, dt,$$

for $f, g \in X$. For each $f \in X$, set

$$I(f) = \int_a^b f(t)\, dt.$$

Prove that the function $I:(X, d^*) \to (R, d)$ is continuous.

2. Let (X_i, d_i), $i = 1, \ldots, n$ be metric spaces. Let $f_i: X_i \to X_{i+1}$, $i = 1, \ldots, n - 1$ be continuous functions. Prove that the function $f_{n-1} \ldots f_1: X_1 \to X_n$ is continuous.

3. Let (X_i, d_i), (Y_i, d_i'), $i = 1, \ldots, n$ be metric spaces. Let $f_i: X_i \to Y_i$, $i = 1, \ldots, n$ be continuous functions. Let

$$X = \prod_{i=1}^{n} X_i \quad \text{and} \quad Y = \prod_{i=1}^{n} Y_i$$

and convert X and Y into metric spaces in the standard manner. Define the function $F: X \to Y$ by

$$F(x_1, x_2, \ldots, x_n) = (f_1(x_1), f_2(x_2), \ldots, f_n(x_n)).$$

Prove that F is continuous.

4. Define the function $f: R^2 \to R$ by

$$f(x_1, x_2) = x_1 + x_2.$$

Prove that f is continuous, where the distance function on R^2 is either d or d'.

5. Define the function $f: R^2 \to R$ by

$$f(x_1, x_2) = x_1 x_2.$$

Prove that f is continuous, where the distance function on R^2 is either d or d'.

6. Let (X, d) be a metric space and $k > 0$. Define the distance function d_k by setting $d_k(x, y) = k \cdot d(x, y)$ for $x, y \in X$. Prove that

$$i: (X, d) \to (X, d_k)$$

and

$$i: (X, d_k) \to (X, d)$$

are continuous, where $i: X \to X$ is the identity function.

7. Let A and B be subsets of R. A function $f: A \to B$ is called *monotone increasing* if $x, y \in A$ and $x < y$ imply $f(x) < f(y)$.
   (a) Let $f: A \to B$ be monotone increasing. Prove that $f: A \to B$ is one-one.
   (b) Let $f: [a, b] \to [f(a), f(b)]$ be monotone increasing and continuous. Prove that this function is invertible.

4 Open Spheres and Neighborhoods

In the definition of continuity of a function f at a point a in a metric space (X, d), we are concerned with how f transforms

those points $x \in X$ such that $d(x, a) < \delta$. If we give a name to this particular collection of points in X we shall be able to cast the definition of continuity in a more terse form.

Definition 4.1 Let (X, d) be a metric space. Let $a \in X$ and $\delta > 0$ be given. The subset of X consisting of those points $x \in X$ such that $d(a, x) < \delta$ is called the *open sphere about a of radius δ* and is denoted by

$$S(a; \delta).$$

Thus, $x \in S(a; \delta)$ if and only if $x \in X$ and $d(x, a) < \delta$. Similarly, if (Y, d') is another metric space and $f: X \to Y$, we have $y \in S(f(a); \varepsilon)$ if and only if $y \in Y$ and $d'(y, f(a)) < \varepsilon$. Thus:

Theorem 4.2 A function $f: (X, d) \to (Y, d')$ is continuous at a point $a \in X$ if and only if given $\varepsilon > 0$ there is a $\delta > 0$ such that

$$f(S(a; \delta)) \subset S(f(a); \varepsilon).$$

For a function $f: X \to Y$ we have $f(A) \subset B$ if and only if $A \subset f^{-1}(B)$, where A and B are subsets of X and Y respectively. Therefore:

Theorem 4.3 A function $f: (X, d) \to (Y, d')$ is continuous at a point $a \in X$ if and only if given $\varepsilon > 0$ there is a $\delta > 0$ such that

$$S(a; \delta) \subset f^{-1}(S(f(a); \varepsilon)).$$

Given a point a in a metric space (X, d), the subset $S(a; \delta)$ of X, for each $\delta > 0$, is an example of the type of subset of X that is called a neighborhood of a.

Definition 4.4 Let (X, d) be a metric space and $a \in X$. A subset N of X is called a *neighborhood of a* if there is a $\delta > 0$ such that

$$S(a; \delta) \subset N.$$

The collection $\mathfrak{N}_a$ of all neighborhoods of a point $a \in X$ is called a *complete system of neighborhoods* of the point a.

A neighborhood of a point $a \in X$ may be thought of as containing all the points of X that are sufficiently close to a or as "enclosing" a by virtue of the fact that it contains some open sphere about a. In particular, for each $\delta > 0$, $S(a; \delta)$ is a neighborhood of a. These open spheres have the property that they are neighborhoods of each of their points.

Lemma 4.5 Let (X, d) be a metric space and $a \in X$. For each $\delta > 0$, the open sphere $S(a; \delta)$ is a neighborhood of each of its points.

Proof. Let $b \in S(a; \delta)$. In order to show that $S(a; \delta)$ is a neighborhood of b we must show that there is an $\eta > 0$ such that

$$S(b; \eta) \subset S(a; \delta).$$

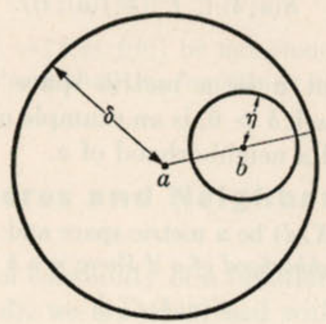
Since $b \in S(a; \delta)$, $d(a, b) < \delta$. Choose $\eta < \delta - d(a, b)$. If $x \in S(b; \eta)$ then

$$d(a, x) \leqq d(a, b) + d(b, x) < d(a, b) + \eta < d(a, b) + \delta - d(a, b) = \delta,$$

and therefore $x \in S(a; \delta)$. Thus $S(b; \eta) \subset S(a; \delta)$ and $S(a; \delta)$ is a neighborhood of b.

We may describe this proof pictorially. We have started with an open sphere $S(a; \delta)$ about a. We choose a point $b \in S(a; \delta)$. Then the minimum distance from b to points not in $S(a; \delta)$ is at least $\delta - d(a, b)$, as indicated in Figure 8, so

Figure 8



that a sphere about b of radius $\eta < \delta - d(a, b)$ is contained in $S(a; \delta)$.

The complete system of neighborhoods of a point is important because this concept may be used to characterize continuity of a function at a point.

Theorem 4.6 Let $f: (X, d) \to (Y, d')$. f is continuous at a point $a \in X$ if and only if for each neighborhood M of $f(a)$ there is a corresponding neighborhood N of a, such that

$$f(N) \subset M,$$

or equivalently,

$$N \subset f^{-1}(M).$$

Proof. First suppose that f is continuous at the point $a \in X$. We must show that, given a neighborhood M of $f(a)$, we can find a neighborhood N of a such that $f(N) \subset M$. Since M is a neighborhood of $f(a)$, there is an $\varepsilon > 0$ such that $S(f(a); \varepsilon) \subset M$. Since f is continuous at a, there is a $\delta > 0$ such that $f(S(a; \delta)) \subset S(f(a); \varepsilon)$. But $N = S(a; \delta)$ is a neighborhood of a, therefore

$$f(N) = f(S(a; \delta)) \subset S(f(a); \varepsilon) \subset M.$$

Conversely, suppose that f satisfies the property that for each neighborhood M of $f(a)$, there is a corresponding neighborhood N of a, such that $f(N) \subset M$. Let $\varepsilon > 0$ be given. To prove that f is continuous at a we must show that there is a $\delta > 0$ such that

$$f(S(a; \delta)) \subset S(f(a); \varepsilon).$$

But $S(f(a); \varepsilon) = M$ is a neighborhood of $f(a)$ and therefore there is a neighborhood N of a such that $f(N) \subset M$. Since N is a neighborhood of a, there is a $\delta > 0$ such that $S(a; \delta) \subset N$. Therefore

$$f(S(a; \delta)) \subset f(N) \subset M = S(f(a); \varepsilon).$$

The proof of the first part of the above theorem may be represented pictorially by considering an arbitrary neighborhood M of $f(a)$ (as indicated in Figure 9). Since M is a neighborhood of $f(a)$, it contains an open sphere $S(f(a); \varepsilon)$ for some $\varepsilon > 0$. Since f is continuous at a, for some $\delta > 0$ the

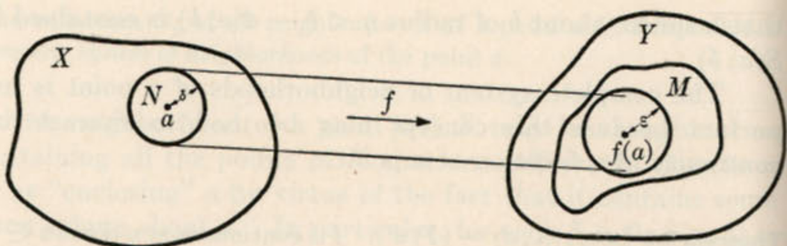


Figure 9

neighborhood $N = S(a; \delta)$ is carried into M by f. Similarly, the proof of the second part of the theorem may be depicted by Figure 10. We start with a neighborhood $M = S(f(a); \varepsilon)$

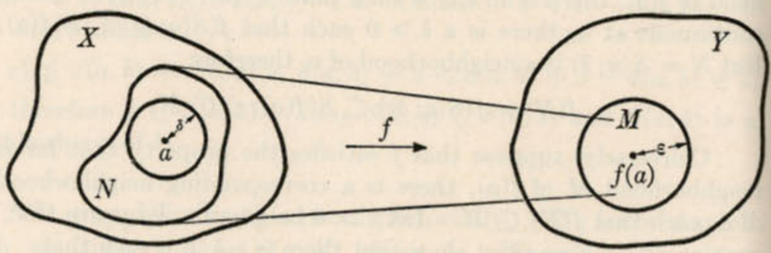


Figure 10

of $f(a)$. The assumed property of f allows us to assert that there is a neighborhood N of a that is carried into M by f. Since N is a neighborhood of a we have an open sphere $S(a; \delta)$ contained in N, which must also be carried into M.

If N is a neighborhood of a point a in a metric space (X, d) and N' is a subset of X that contains N, then N' contains the same open sphere about a that N does and therefore N' is also a neighborhood of a. Thus, the previous theorem becomes:

Theorem 4.7 Let $f: (X, d) \to (Y, d')$. f is continuous at a point $a \in X$ if and only if for each neighborhood M of $f(a)$, $f^{-1}(M)$ is a neighborhood of a.

The characterization of continuity in terms of neighborhoods provides a relatively simple method for proving many theorems involving continuity. For example, had we postponed the proof of the theorem that the composition of two continuous functions is a continuous function, we could now prove this theorem in the following manner:

Theorem Let $f: X \to Y$ be continuous at $a \in X$ and $g: Y \to Z$ be continuous at $f(a) \in Y$, then $gf: X \to Z$ is continuous at $a \in X$.

Proof. Let M be a neighborhood of $g(f(a))$. Then $g^{-1}(M)$ is a neighborhood of $f(a)$ and $f^{-1}(g^{-1}(M))$ is a neighborhood of a. Since $f^{-1}(g^{-1}(M)) = (gf)^{-1}(M)$, the function gf is continuous at a.

As another application of the concept of neighborhood we have:

Theorem 4.8 Let $f: R \to R$ be continuous at $a \in R$ and let $f(a) > 0$. Then there is a $\delta > 0$ such that $f(x) > \frac{f(a)}{2}$ whenever $|x - a| < \delta$.

Proof. The open interval $M = \left(\frac{f(a)}{2}, \frac{3f(a)}{2}\right)$ is a neighborhood of $f(a)$, therefore $f^{-1}(M)$ is a neighborhood of a. Consequently, there is a $\delta > 0$ such that $(a - \delta, a + \delta) \subset f^{-1}(M)$ or $f(x) > \frac{f(a)}{2}$ for $|x - a| < \delta$.

The collections of neighborhoods of points in a metric space possess five properties that will be of significance in the next chapter.

Theorem 4.9 Let (X, d) be a metric space.

*N*0. For each point $a \in X$, there exists at least one neighborhood of a.

$N1$. For each point $a \in X$ and each neighborhood N of a, $a \in N$.

$N2$. For each point $a \in X$, if N is a neighborhood of a and $N' \supset N$, then N' is a neighborhood of a.

$N3$. For each point $a \in X$ and each pair N, M of neighborhoods of a, $N \cap M$ is also a neighborhood of a.

$N4$. For each point $a \in X$ and each neighborhood N of a, there exists a neighborhood O of a such that $O \subset N$ and O is a neighborhood of each of its points.

Proof. For $a \in X$, X is a neighborhood of a, thus $N0$ is true. $N1$ is trivial and $N2$ has already been discussed. To prove $N3$, let N and M be neighborhoods of a. Then N and M contain open spheres $S(a; \delta_1)$ and $S(a; \delta_2)$ respectively and therefore $N \cap M$ contains the open sphere $S(a; \delta)$ where $\delta = \text{minimum } \{\delta_1, \delta_2\}$. To prove $N4$, let N be a neighborhood of a. Then N contains an open sphere $S(a; \delta)$ and by lemma 4.5, $O = S(a; \delta)$ is a neighborhood of each of its points.

There are other properties of the neighborhoods of points in a metric space that often play an important role in topological considerations.

Theorem 4.10 Let (X, d) be a metric space.

$AC1$ (First axiom of countability). For each point $a \in X$, there is a sequence $O_1, O_2, \ldots, O_n, \ldots$ of neighborhoods of a with the property that each neighborhood N of a contains at least one neighborhood of this sequence.

$T2$ (Hausdorff axiom). For each pair x, y of distinct points of X, there is a neighborhood M of x and a neighborhood N of y such that $M \cap N = \emptyset$.

Proof. To prove the first axiom of countability we simply exhibit the sequence of neighborhoods of a point a in which the n^{th} neighborhood $O_n = S\left(a; \frac{1}{n}\right)$. To prove the Hausdorff axiom, we have, for $a \neq b$, $d(a, b) > 0$. If we choose $\delta = \frac{d(a, b)}{3}$, then the intersection of the neighborhoods $S(a; \delta)$ and $S(b; \delta)$ of a and b respectively is the empty set.

Exercises

1. Let $a \in R$. Prove that a subset M of R is a neighborhood of a if and only if M contains an open interval containing a.
2. Let $O = O_1 \cup O_2 \cup \ldots \cup O_n \cup \ldots$, where for each positive integer i, O_i is an open interval. Prove that O is a neighborhood of each of its points.
3. Let $O = O_1 \cap O_2 \cap \ldots \cap O_n$, where for $i = 1, \ldots, n$, each O_i is an open interval. Prove that O is a neighborhood of each of its points.
4. For each positive integer i, let $O_i = \left(-\frac{1}{i}, 1 + \frac{1}{i}\right)$. Prove that $O = \bigcap_{i=1}^{\infty} O_i = [0, 1]$ and the O is not a neighborhood of each of its points.
5. Let $a \in R$ and $f: R \to R$ be defined by $f(x) = 0$ for $x \leqq a$, $f(x) = 1$ for $x > a$. Prove that f is not continuous at a, but is continuous at all other points.
6. Let (X, d) be a metric space such that $d(x, y) = 1$ whenever $x \neq y$. Let $a \in X$. Prove that $\{a\}$ is a neighborhood of a. Prove that a subset N of X is a neighborhood of a if and only if $a \in N$. Let O be a subset of X. Prove that O is a neighborhood of each of its points.
7. Let $(X_1, d_1), (X_2, d_2), \ldots, (X_n, d_n)$ be metric spaces and convert
$$X = \prod_{i=1}^{n} X_i$$
into a metric space (X, d) in the standard manner. Prove that an open sphere in (X, d) is the product of open spheres from $X_1, X_2, \ldots, X_n$. Prove that the product of open spheres from $X_1, X_2, \ldots, X_n$ contains an open sphere of X.
8. In (R^2, d), where d is defined as in Section 2, prove that for each point $(a_1, a_2) \in R^2$, the subset $O_1 \times O_2$, where O_1 is an open interval containing a_1 and O_2 is an open interval containing a_2, is a neighborhood of (a_1, a_2). Prove that a subset N of R^2 is a neighborhood of (a_1, a_2) if and only if N contains a set of the form $O_1 \times O_2$ as above.

5 Open Sets

In a metric space, the open sphere $S(a; \delta)$ is a neighborhood of each of its points (Lemma 4.5). The collection of subsets possessing this property plays a fundamental role in topology.

Definition 5.1 A subset O of a metric space is said to be *open* if O is a neighborhood of each of its points.

In particular, each open sphere $S(a; \delta)$ is open. Open sets may be characterized directly in terms of open spheres.

Theorem 5.2 A subset O of a metric space (X, d) is an open set if and only if it is a union of open spheres.

Proof. Suppose O is open. Then for each $a \in O$, there is an open sphere $S(a; \delta_a) \subset O$. Therefore

$$O = \bigcup_{a \in O} S(a; \delta_a)$$

is a union of open spheres. Conversely, if O is a union of open spheres, then using the centers of these spheres as the elements of an indexing set we can write

$$O = \bigcup_{a \in I} S(a; \delta_a).$$

If $x \in O$, then $x \in S(a; \delta_a)$ for some $a \in I$. $S(a; \delta_a)$ is a neighborhood of x and since $S(a; \delta_a) \subset O$, by $N2$, O is a neighborhood of x. Thus O is a neighborhood of each of its points, and by Definition 5.1, O is open.

Most of the functions considered in topology are continuous. Open sets provide a simple characterization of continuity.

Theorem 5.3 Let $f: (X, d) \to (Y, d')$. Then f is continuous if and only if for each open set O of Y, the subset $f^{-1}(O)$ is an open subset of X.

Proof. First, suppose f is continuous. Let $O \subset Y$ be open. We must show that $f^{-1}(O)$ is open; that is, $f^{-1}(O)$ is a neighborhood of each of its points. To this end, let $a \in f^{-1}(O)$, then $f(a) \in O$ and O is a neighborhood of $f(a)$. Since f is continuous at a, Theorem 4.7 may be applied, yielding $f^{-1}(O)$ is a neighborhood of a.

Conversely, suppose for each open set $O \subset Y$, $f^{-1}(O)$ is open. Let $a \in X$ and let M be a neighborhood of $f(a)$. Then there is an $\varepsilon > 0$ such that $S(f(a); \varepsilon) \subset M$. But $S(f(a); \varepsilon)$ is open and therefore $f^{-1}(S(f(a); \varepsilon))$ is open. Since $a \in f^{-1}(S(f(a); \varepsilon))$, this subset is a neighborhood of a. Therefore $f^{-1}(M)$ contains a neighborhood of a and f is continuous at a. Since a was arbitrary, f is continuous.

Theorem 5.4 Let a be a point of a metric space (X, d). Then a subset N of X is a neighborhood of a if and only if N contains an open set that in turn contains a.

Proof. If N is a neighborhood of a, then N contains an open set, namely an open sphere with center at a, which in turn contains a. Conversely, given $a \in O \subset N$, where O is an open set, O is necessarily a neighborhood of a, and therefore by $N2$, N is a neighborhood of a.

The adjectives "trivial" or "obvious" could well be applied to Theorem 5.4, and this theorem would certainly not be worth stating were it not that this "trivial" statement will have important consequences when we attempt to extend the concepts of neighborhood and open set to more general situations. The form of Theorem 5.4 indicates that the term *neighborhood* may be defined by means of the term *open set*. By this is meant a statement of the form "N is a neighborhood of a if and only if N has some property with respect to open sets." An immediate consequence of the form of this statement is that there is an alternate organization of the material in Sections 4 and 5. We could have first defined certain sets to be open sets directly. Let us call them "tentatively open" for the moment.

Definition (alternate). A subset O of a metric space is called tentatively open if it is a union of open spheres.

We could then follow this alternate definition by:

Definition (alternate). A subset N of a metric space is called a tentative neighborhood of a point a if N contains a tentatively open set that contains a.

Note that the form of the first of these two definitions is dictated by the form of Theorem 5.2, and the form of the second definition is dictated by the form of Theorem 5.4. Theorem 5.2 then asserts that a subset O of a metric space is open if and only if it is tentatively open and, consequently, Theorem 5.4 asserts that a subset N of a metric space is a neighborhood of a point a if and only if it is a tentative neighborhood of a. We can therefore say that, had we chosen to follow the scheme of organization indicated by these two definitions, we could delete the adjective tentative, for we obtain precisely the same neighborhoods and open sets in a given metric space in this manner as the neighborhoods and open sets that we have obtained by using Definitions 4.4 and 5.1.

Just as the collections of neighborhoods of points in a metric space possess certain significant properties that it is desirable to record, so do the collection of open sets in a metric space.

Theorem 5.5 Let (X, d) be a metric space.

$O1$. The empty set is open.

$O2$. X is open.

$O3$. If $O_1, O_2, \ldots, O_n$ are open, then

$$O_1 \cap O_2 \cap \cdots \cap O_n$$

is open.

$O4$. If for each $\alpha \in I$, O_α is an open set, then

$$\cup_{\alpha \in I} O_\alpha$$

is open.

Proof. The empty set is open, for in order for it not to be open there would have to be a point $x \in \emptyset$. Given a point $a \in X$, for any $\delta > 0$, $S(a; \delta) \subset X$ and therefore X is a neighborhood of each of its points; that is, X is open. To prove $O3$, let

$$a \in O_1 \cap O_2 \cap \cdots \cap O_n,$$

where for $i = 1, 2, \ldots, n$, O_i is open. Then each O_i is a neighborhood of a. By $N3$, the intersection of two neighborhoods of a is again a neighborhood of a, and hence by induction, the intersection of a finite number of neighborhoods of a is again a neighborhood of a. Therefore $O_1 \cap O_2 \cap \cdots \cap O_n$ is a neighborhood of each of its points. Finally, to prove $O4$, let

$$a \in O = \cup_{\alpha \in I} O_\alpha,$$

where for each $\alpha \in I$, O_α is open. Then $a \in O_\beta$ for some $\beta \in I$ and O_β is a neighborhood of a. Since $O_\beta \subset O$, by $N2$, O is a neighborhood of a. Therefore O is a neighborhood of each of its points.

Exercises

1. Let (X_i, d_i), $i = 1, 2, \ldots, n$ be metric spaces. Let

$$X = \prod_{i=1}^{n} X_i$$

and let (X, d) be the metric space defined in the standard manner by Theorem 2.3. For $i = 1, 2, \ldots, n$, let O_i be an open subset of X_i. Prove that the subset $O_1 \times O_2 \times \cdots \times O_n$ of X is open and that each open subset of X is a union of sets of this form. [A collection of open sets of a metric space is called a *basis for the open sets* if each open set is a union of sets in this collection. For example, the open spheres in a metric space form a basis for the open sets.]

2. Let X be a set and d the distance function on X defined by $d(x, x) = 0$, $d(x, y) = 1$ for $x \neq y$. Prove that each subset of (X, d) is open.

3. Let (R^n, d), (R^n, d'), (R^n, d'') be as in Section 2. Prove that the following three statements are equivalent:

1. O is open in (R^n, d).
2. O is open in (R^n, d').
3. O is open in (R^n, d'').

4. Let (X, d) be a metric space, let $k > 0$, and let d_k be the distance function defined by $d_k(x, y) = k \cdot d(x, y)$, $x, y \in X$. Prove that a subset O of X is open in (X, d) if and only if it is open in (X, d_k).

6 Limit Points

In a metric space, the topological concepts of continuity, neighborhood, and open set are often defined in terms of "limits of sequences." We shall first describe this new concept as applied to the real number system.

Definition 6.1 Let $a_1, a_2, \ldots$ be a sequence of real numbers. A real number a is said to be the *limit of the sequence* $a_1, a_2, \ldots$ if, given $\varepsilon > 0$, there is a positive integer N such that, whenever $n > N$,

$$|a - a_n| < \varepsilon.$$

In this event we shall also say that the sequence $a_1, a_2, \ldots$ *converges to* a and write

$$\lim_n a_n = a.$$

Interpreting ε as an "arbitrary degree of closeness" and N as "sufficiently far out in the sequence," we see that we have defined $\lim_n a_n = a$ in the event that a_n may be made arbitrarily close to a by requiring that a_n be sufficiently far out in the sequence.

Now, suppose that we have a metric space (X, d) and a sequence $a_1, a_2, \ldots$ of points of X. Given a point $a \in X$ we measure the distance from a to the successive points of the sequence, by the sequence of real numbers $d(a, a_1), d(a, a_2), \ldots$. It is natural to say that the limit of the sequence $a_1, a_2, \ldots$ of points of X is the point a if the limit of the sequence of real numbers $d(a, a_1), d(a, a_2), \ldots$ is the real number 0.

Definition 6.2 Let (X, d) be a metric space. Let $a_1, a_2, \ldots$ be a sequence of points of X. A point $a \in X$ is said to be the *limit of the sequence* $a_1, a_2, \ldots$ if

$$\lim_n d(a, a_n) = 0.$$

Again, in this event, we shall say that the sequence $a_1, a_2, \ldots$ *converges to* a and write

$$\lim_n a_n = a.$$

The statement, "$\lim_n d(a, a_n) = 0$," is by Definition 6.1 an abbreviation of the statement, "given $\varepsilon > 0$, there is an integer N such that, whenever $n > N$, $d(a, a_n) < \varepsilon$," or equivalently $a_n \in S(a; \varepsilon)$. In the event that $(X, d) = (R, d)$, Definitions 6.1 and 6.2 agree. Furthermore,

Lemma 6.3 Let (X, d) be a metric space and $a_1, a_2, \ldots$ be a sequence of points of X. Then $\lim_n a_n = a$ for a point $a \in X$ if and only if, given $\varepsilon > 0$, there is an integer N, such that, whenever $n > N$, then

$$a_n \in S(a; \varepsilon).$$

This lemma could be used as a substitute for the two previous definitions. It also allows us to characterize limits of sequences in terms of either neighborhoods or open sets.

Theorem 6.4 Let (X, d) be a metric space and $a_1, a_2, \ldots$ be a sequence of points of X. Then $\lim_n a_n = a$ for a point $a \in X$ if and only if for each neighborhood V of a there is an integer N such that $a_n \in V$ whenever $n > N$.

Theorem 6.5 Let (X, d) be a metric space and $a_1, a_2, \ldots$ be a sequence of points of X. Then $\lim_n a_n = a$ for a point $a \in X$ if and only if for each open set O that contains a there is an integer N such that $a_n \in O$ whenever $n > N$.

The proofs of these two theorems are straightforward applications of Lemma 6.3 and are left as exercises.

In Section 4 we proved that a metric space satisfies the Hausdorff axiom; that is, if a and b are distinct points, then there are neighborhoods U and V of a and b respectively, such that $U \cap V = \emptyset$. From this result we obtain the uniqueness of limits.

Theorem 6.6 In a metric space, let $\lim_n a_n = a$ and $\lim_n a_n = b$, then $a = b$.

Proof. Suppose $a \neq b$. Then there are neighborhoods U and V of a and b respectively, such that $U \cap V = \emptyset$. Thus, if $\lim_n a_n = a$, we have, for some integer N, $a_n \in U$ whenever $n > N$ and, consequently, $a_n \notin V$ for $n > N$. Therefore $\lim_n a_n = b$ is impossible.

Continuity may be characterized in terms of limits of sequences in accordance with the following theorem.

Theorem 6.7 Let (X, d), (Y, d') be metric spaces. A function $f: X \to Y$ is continuous at a point $a \in X$ if and only if, whenever $\lim_n a_n = a$ for a sequence $a_1, a_2, \ldots$ of points of X, $\lim_n f(a_n) = f(a)$.

Proof. Suppose f is continuous at a and $\lim_n a_n = a$. Let V be a neighborhood of $f(a)$. Then $f^{-1}(V)$ is a neighborhood of a, so by Theorem 6.4 there is an integer N such that $a_n \in f^{-1}(V)$ whenever $n > N$. Consequently, $f(a_n) \in V$ whenever $n > N$. Thus, for each neighborhood V of $f(a)$ there is an integer N such that $f(a_n) \in V$ whenever $n > N$ and again, applying Theorem 6.4, $\lim_n f(a_n) = f(a)$.

To prove the "if" part of this theorem, we shall prove that if f is not continuous at a, then there is at least one sequence $a_1, a_2, \ldots$ of points of X, such that $\lim_n a_n = a$, but $\lim_n f(a_n) = f(a)$ is false. Since f is not continuous at a, there is a neighborhood V of $f(a)$ such that for each neighborhood U of a,

$$f(U) \not\subset V.$$

In particular, for each neighborhood $S\left(a; \frac{1}{n}\right)$, $n = 1, 2, \ldots$

$$f\left(S\left(a; \frac{1}{n}\right)\right) \not\subset V.$$

Thus, for each positive integer n, there is a point a_n with $a_n \in S\left(a; \frac{1}{n}\right)$ and $f(a_n) \notin V$. Now $d(a, a_n) < \frac{1}{n}$ and therefore $\lim_n a_n = a$, whereas, $\lim_n f(a_n) = f(a)$ is impossible, since $f(a_n) \notin V$ for all n.

If $\lim_n a_n = a$, we can write $\lim_n f(a_n) = f(a)$ as $\lim_n f(a_n) = f(\lim_n a_n)$. We may therefore describe a continuous function as one that commutes with the operation of taking limits.

Exercises

1. Each of the points $a_1, a_2, \ldots$ of a sequence of points of R^k has k coordinates; that is,
$$a_n = (a_1^n, a_2^n, \ldots, a_k^n) \in R^k, n = 1, 2, \ldots.$$
Let
$$c = (c_1, c_2, \ldots, c_k) \in R^k.$$
In each of the three metric spaces (R^k, d), (R^k, d'), (R^k, d''), prove that $\lim_n a_n = c$ if and only if
$$\lim_n a_i^n = c_i, i = 1, 2, \ldots, k.$$

2. Let (X_i, d_i), $i = 1, 2, \ldots, n$ be metric spaces. Let $X = \prod_{i=1}^{n} X_i$ and let (X, d) be the metric space defined in the standard manner by Theorem 2.3. For $i = 1, 2, \ldots, n$ define the i^{th} projection $p_i: X \to X_i$ by $p_i(x_1, x_2, \ldots, x_n) = x_i$. Prove that each p_i is continuous. Let (Y, d') be a metric space. Prove that a function $f: (Y, d') \to (X, d)$ is continuous if and only if the n functions $p_i f: (Y, d') \to (X_i, d_i)$, $i = 1, 2, \ldots, n$ are continuous.

3. Let $a_1, a_2, \ldots$ be a sequence of real numbers, each of which is in a certain closed interval $[c, d]$. Prove that if $\lim_n a_n = a$, then $a \in [c, d]$.

4. Define the concept of subsequence and prove that in a metric space a subsequence of a convergent sequence is convergent and has the same limit as the original sequence.

5. Let $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ be two sequences of points of a metric space such that $a_i = b_i$ for all but a finite number of positive integers i. Prove that $\lim_n a_n = a$ if and only if $\lim_n b_n = a$.

7 Closed Sets

Definition 7.1 A subset F of a metric space is said to be *closed* if its complement, $C(F)$, is open.

In the real number system, a closed interval $[a, b]$ is a closed set, for its complement is the union of the two open sets O_1 and O_2, where O_1 is the set of real numbers x such that $x < a$ and O_2 is the set of real numbers x such that $x > b$. A common mistake is the assumption that a set cannot be both open and closed. In any metric space (X, d), the two sets $\emptyset$ and X are open, and therefore their complements X and $\emptyset$ are closed. Thus, X and also $\emptyset$ are both open and both closed. Whether or not, in a given metric space, there are other subsets that are simultaneously open and closed, is a significant topological property, which we shall subsequently describe by the adjective "connected." In any event, the adjectives *open* and *closed* are not mutually exclusive. Nor, for that matter, are they all-inclusive, for we shall shortly give an example of a subset of the real number system that is neither open nor closed.

In a metric space, a closed set may be described as a set that contains all its limit points; that is:

Theorem 7.2 In a metric space (X, d), a set $F \subset X$ is closed if and only if for each sequence $a_1, a_2, \ldots$ of points of F that converges to a point $a \in X$ we have $a \in F$.

Proof. First, let F be closed. Suppose $\lim_n a_n = a$ and $a_n \in F$ for $n = 1, 2, \ldots$. We shall show that the assumption $a \in C(F)$ leads to a contradiction. $C(F)$ is open, therefore $a \in C(F)$ and $\lim_n a_n = a$ implies that there is an integer N such that for $n > N$, $a_n \in C(F)$, which contradicts the hypothesis $a_n \in F$ for $n = 1, 2, \ldots$. Conversely, suppose that F is a set such that for each sequence with $\lim_n a_n = a$ and $a_n \in F$ for $n = 1, 2, \ldots$ we have $a \in F$. We must show that F is closed, or, equivalently, that $C(F)$ is open. If $C(F)$ is not open, then there must be a point $c \in C(F)$ such that for each neighborhood N of c,

$$N \not\subset C(F).$$

In particular,

$$S\left(c; \frac{1}{n}\right) \not\subset C(F)$$

for $n = 1, 2, \ldots$. Thus, there is a sequence of points $a_1, a_2, \ldots$ such that $a_n \in F$ for $n = 1, 2, \ldots$ and $\lim_n a_n = c$. The assumed property of the set F then leads to the contradiction that $c \in F$. Therefore, $C(F)$ is open and F is closed.

Theorem 7.2 might be viewed more directly as a proof of the statement

Corollary 7.3 In a metric space (X, d), a set $O \subset X$ is open if and only if for each sequence $a_1, a_2, \ldots$ of points of $C(O)$ that converges to a point $a \in X$ we have $a \notin O$.

In keeping with the theme of the last three sections we should anticipate a theorem characterizing continuity in terms of closed sets.

Theorem 7.4 Let (X, d), (Y, d') be metric spaces. A function $f: X \to Y$ is continuous if and only if for each closed subset A of Y, the set $f^{-1}(A)$ is a closed subset of X.

Proof. For $A \subset Y$, we have

$$C(f^{-1}(A)) = f^{-1}(C(A)).$$

But f is continuous if and only if the inverse image of each open set is an open set, and this is true if and only if the inverse image of each closed set is a closed set.

Theorem 7.5 Let (X, d) be a metric space.

$C1$. X is closed.

$C2$. $\emptyset$ is closed.

$C3$. The union of a finite collection of closed sets is closed.

$C4$. The intersection of a family of closed sets is closed.

Proof. $C1$ and $C2$ have already been discussed. $C3$ and $C4$ follow from the application of DeMorgan's formulas to the corresponding properties $O3$ and $O4$ of open sets.

The union of closed sets need not, in general, be a closed set, as may be seen by the following example. For each positive integer n let F_n be the closed interval $\left[\frac{1}{n}, 1\right]$. Then $\bigcup_{n=1}^{\infty} F_n = (0, 1]$, where $(0, 1]$ is the set of real numbers x such that $0 < x \leqq 1$. The set $(0, 1]$ is not closed for the sequence of points $a_1, a_2, \ldots$ where $a_n = \frac{1}{n}$ is such that $a_n \in (0, 1]$ for each n, whereas $\lim_n a_n = 0$, and $0 \notin (0, 1]$ (Theorem 7.2). For that matter $(0, 1]$ is not open, since the sequence of points $b_1, b_2, \ldots$ where $b_n = 1 + \frac{1}{n}$ is such that $b_n \notin (0, 1]$ for each n, whereas $\lim_n b_n = 1$ and $1 \in (0, 1]$ (Corollary 7.3).

We shall conclude this section by presenting a characterization of closed sets in terms of the distance function. This characterization utilizes the concept of the distance between a point and a subset, which, in turn, requires the result that a non-empty set of real numbers that has a lower bound has a greatest lower bound.

Definition 7.6 Let A be a set of real numbers. A real number b is called a *lower bound* of A if $b \leqq x$ for each $x \in A$. A lower bound b^* of A is called a *greatest lower bound* (g.l.b.) of A if for each lower bound b of A, $b \leqq b^*$.

The greatest lower bound of a set A of real numbers may or may not be an element of A. For example, 0 is a g.l.b. of $[0, 1]$ and $0 \in [0, 1]$, whereas 0 is also a g.l.b. of $(0, 1)$ but $0 \notin (0, 1)$. In any event, the following is true:

Theorem 7.7 A non-empty subset A of real numbers that has a lower bound has a greatest lower bound.

Proof. Define the set B by $x \in B$ if and only if $-x \in A$. Then B is non-empty and the negative of a lower bound of A is an upper bound of B. Since the set of real numbers is complete, B has a least upper bound. But the negative of a least upper bound of B is a greatest lower bound of A.

Lemma 7.8 Let b be a greatest lower bound of the non-empty subset A. Then, for each $\varepsilon > 0$, there is an element $x \in A$ such that

$$x - b < \varepsilon.$$

Proof. Suppose there were an $\varepsilon > 0$ such that $x - b \geqq \varepsilon$ for each $x \in A$. Then $b + \varepsilon \leqq x$ for each $x \in A$ and $b + \varepsilon$ would be a lower bound of A. Since b is a g.l.b. of A, we obtain the contradiction $b + \varepsilon \leqq b$.

Corollary 7.9 Let b be a greatest lower bound of the non-empty subset A of real numbers. Then there is a sequence $a_1, a_2, \ldots$ of real numbers such that $a_n \in A$ for each n and $\lim_n a_n = b$.

Proof. For $\varepsilon = \frac{1}{n}$ we obtain an element $a_n \in A$ such that

$$a_n - b < \frac{1}{n}.$$

Since b is a lower bound of A, $0 \leqq a_n - b$. Therefore $\lim_n a_n = b$.

Definition 7.10 Let (X, d) be a metric space. Let $a \in X$ and let A be a non-empty subset of X. The greatest lower bound of the set of numbers of the form $d(a, x)$ for $x \in A$ is called the *distance between a and A* and is denoted by

$$d(a, A).$$

From Corollary 7.9 we obtain

Corollary 7.11 Let (X, d) be a metric space, $a \in X$, and A a non-empty subset of X. Then there is a sequence $a_1, a_2, \ldots$ of points of A such that $\lim_n d(a, a_n) = d(a, A)$.

Theorem 7.12 A subset F of a metric space (X, d) is closed if and only if for each point $x \in X$, $d(x, F) = 0$ implies $x \in F$.

Proof. First, suppose F is closed. Let $x \in X$ be such that $d(x, F) = 0$. By Theorem 7.11 there is a sequence of points of F such that $\lim_n d(x, a_n) = 0$. Thus, $\lim_n a_n = x$, and hence by 7.2, $x \in F$. Conversely, suppose that F is such that $d(x, F) = 0$ implies $x \in F$. Let $a_1, a_2, \ldots$ be a sequence of points of F such that $\lim_n a_n = x$. To show that F is closed we must show that $x \in F$. In view of the condition imposed on F, it suffices to show that $d(x, F) = 0$. But, given $\varepsilon > 0$, there is an integer n such that $d(x, a_n) < \varepsilon$. Since $a_n \in F$, $d(x, F) < \varepsilon$ for each $\varepsilon > 0$. Thus, $d(x, F) = 0$ and $x \in F$.

Exercises

1. Let (X, d_1), (Y, d_2) be metric spaces. Let $f: X \to Y$ be continuous. Define a distance function d on $X \times Y$ in the standard manner. Prove that the graph Γ_f of f is a closed subset of $(X \times Y, d)$.
2. Let $F: R \to R$ be defined by
$$f(x) = \frac{1}{x}, \ x > 0,$$
$$f(x) = 0, \ x \leqq 0.$$
Prove that the graph Γ_f is a closed subset of (R^2, d), but that f is not continuous.
3. Let A be a non-empty subset of a metric space (X, d). Define the function $f: X \to R$ by $f(x) = d(x, A)$. Prove that f is continuous.
4. Let (X, d) be a metric space and A a non-empty subset of X. For $x, y \in X$, prove that
$$d(x, A) \leqq d(x, y) + d(y, A).$$

5. Let A be a non-empty subset of a metric space (X, d) and let $x \in X$. Prove that $d(x, A) = 0$ if and only if every neighborhood of x contains a point of A.
6. Let A be a closed, non-empty subset of the real numbers that has a lower bound. Prove that A contains its greatest lower bound.

8 Products

We have indicated that there is a standard procedure for converting into a metric space the direct product of the underlying sets of a finite number of metric spaces. We shall now examine systematically what the neighborhoods, open sets, limits of sequences, and closed sets are in such a space. Throughout this section, then, we shall let $(X_1, d_1), (X_2, d_2), \ldots, (X_n, d_n)$ be metric spaces, $X = \prod_{i=1}^{n} X_i$, and d the distance function defined on X by

$$d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \underset{1 \leq i \leq n}{\text{maximum}} \{d_i(x_i, y_i)\}.$$

The following lemma is a direct consequence of the definition of d as the maximum distance between "coordinates."

Lemma 8.1 Let $a = (a_1, a_2, \ldots, a_n) \in X$, $\delta > 0$. Then

$$S(a; \delta) = S(a_1; \delta) \times S(a_2; \delta) \times \ldots \times S(a_n; \delta).$$

Theorem 8.2 Given a point $a = (a_1, a_2, \ldots, a_n) \in X$, a subset $W \subset X$ is a neighborhood of a if and only if W contains a set of the form

$$N_1 \times N_2 \times \ldots \times N_n,$$

where each N_i is a neighborhood of a_i.

Proof. First, suppose that W contains a set of the form $N_1 \times N_2 \times \ldots \times N_n$, where each N_i is a neighborhood of a_i. Then, for each i, we have a $\delta_i > 0$ such that $S(a_i; \delta_i) \subset N_i$. Let

$$\delta = \text{minimum } \{\delta_1, \delta_2, \ldots, \delta_n\}.$$

Then

$$S(a; \delta) \subset S(a_1; \delta_1) \times S(a_2; \delta_2) \times \ldots \times S(a_n; \delta_n) \subset N_1 \times N_2 \times \ldots \times N_n \subset W.$$

But, $S(a; \delta)$ is a neighborhood of a and, consequently, so is W.

Conversely, suppose W is a neighborhood of a. Then for some $\delta > 0$, $S(a; \delta) \subset W$. If we set $N_i = S(a_i; \delta)$, $i = 1, 2, \ldots, n$, we have

$$S(a; \delta) = N_1 \times N_2 \times \ldots \times N_n \subset W.$$

We have now described the neighborhoods in the product space X in terms of the neighborhoods in each X_i. The word "basis" is often used in this situation.

Definition 8.3 Let a be a point in a metric space. A collection $\mathfrak{N}$ of subsets is called a *basis for the neighborhoods at the point a* if:

1. Each subset in $\mathfrak{N}$ is a neighborhood of a;
2. A subset U is a neighborhood of a if and only if it contains an element of $\mathfrak{N}$.

Theorem 8.2 thus becomes:

Corollary 8.4 A basis for the neighborhoods at a point

$$a = (a_1, a_2, \ldots, a_n) \in X$$

is the collection of subsets of X of the form

$$N_1 \times N_2 \times \ldots \times N_n,$$

where each N_i is a neighborhood of a_i.

We shall next describe the open sets in X.

Theorem 8.5 A subset O of X is open if and only if O is the union of sets of the form

$$O_1 \times O_2 \times \ldots \times O_n,$$

where each O_i is an open subset of X_i.

Proof. We first note that if, for $i = 1, 2, \ldots, n$, O_i is an open subset of X_i, then by Theorem 8.2, $O_1 \times O_2 \times \ldots \times O_n$ is a neighborhood of each of its points and is therefore open. Now, suppose that for each $\alpha \in I$, we have open subsets $O_{\alpha,1}$ of X_1, $O_{\alpha,2}$ of $X_2, \ldots, O_{\alpha,n}$ of X_n. Then, for each $\alpha \in I$,

$$O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n}$$

is an open subset of X. Since the union of open sets is open

$$\bigcup_{\alpha\in I} O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n}$$

is an open subset of X.

Conversely, suppose O is an open subset of X. For each $a = (a_1, a_2, \ldots, a_n) \in O$, there is an open sphere

$$S(a; \delta_a) = S(a_1; \delta_a) \times S(a_2; \delta_a) \times \ldots \times S(a_n; \delta_a) \subset O.$$

Therefore

$$O = \bigcup_{a\in O} S(a_1; \delta_a) \times S(a_2; \delta_a) \times \ldots \times S(a_n; \delta_a).$$

Again, in the following sense, we have used the open subsets of the X_i's to describe a "basis" of the open subsets of X.

Definition 8.6 Let (Y, d') be a metric space and let $\mathfrak{B}$ be a collection of subsets of Y. The collection $\mathfrak{B}$ is called a *basis for the open sets of* (Y, d'), provided that:

1. Each subset in $\mathfrak{B}$ is an open set;
2. A subset V of Y is open if and only if V is the union of sets belonging to the collection $\mathfrak{B}$.

Theorem 8.5 may now be restated:

Corollary 8.7 A basis for the open sets of (X, d) is the collection of subsets of X of the form

$$O_1 \times O_2 \times \ldots \times O_n,$$

where each O_i is an open subset of X_i.

Theorem 8.8 Let $a_1, a_2, \ldots$ be a sequence of points of X. Let $a_j = (a_1^j, a_2^j, \ldots, a_n^j)$ and let $c = (c_1, c_2, \ldots, c_n) \in X$. Then

$$\lim_j a_j = c$$

if and only if, for $i = 1, 2, \ldots, n$,

$$\lim_j a_i^j = c_i.$$

Proof. First suppose that $\lim_j a_j = c$. Let U_i be a neighborhood of c_i, $i = 1, 2, \ldots, n$. To prove that $\lim_j a_i^j = c_i$, it suffices to prove that there is an integer N such that for $j > N$, $a_i^j \in U_i$. By Theorem 8.2, $U_1 \times U_2 \times \ldots \times U_n$ is a neighborhood of c. Since $\lim_j a_j = c$, there is an integer N such that, for $j > N$, $a_j \in U_1 \times U_2 \times \ldots \times U_n$, which, in turn, yields $a_i^j \in U_i$ for $j > N$ and $i = 1, 2, \ldots, n$.

Conversely, suppose that $\lim_j a_i^j = c_i$ for $i = 1, 2, \ldots, n$. Let W be a neighborhood of c. By Theorem 8.2, W contains a set of the form $U_1 \times U_2 \times \ldots \times U_n$, where each U_i is a neighborhood of c_i. Since $\lim_j a_i^j = c_i$, for each i, there is an integer N_i, such that $a_i^j \in U_i$ whenever $j > N_i$. Let $N = \text{maximum } \{N_1, N_2, \ldots, N_n\}$. Then, for $j > N$, $a_j = (a_1^j, a_2^j, \ldots, a_n^j) \in N_1 \times N_2 \times \ldots \times N_n \subset W$. Thus, for each neighborhood W of c there is an integer N such that $a_j \in W$ for $j > N$, and, therefore, $\lim_j a_j = c$.

With regard to closed sets in a product space, let us first make the observation that, if for $i = 1, 2, \ldots, n$, $A_i \subset X_i$, then

$$\begin{aligned} &C(A_1 \times A_2 \times \ldots \times A_n) \\ &= (C(A_1) \times X_2 \times \ldots \times X_n) \cup (X_1 \times C(A_2) \times \ldots \times X_n) \\ &\qquad \cup \ldots \cup (X_1 \times X_2 \times \ldots \times C(A_n)). \end{aligned}$$

Thus, if F_i is a closed subset of X_i for $i = 1, 2, \ldots, n$,

$$F_1 \times F_2 \times \ldots \times F_n$$

is a closed subset of X, for

$$\begin{aligned} &C(F_1 \times F_2 \times \ldots \times F_n) \\ &= (C(F_1) \times X_2 \times \ldots \times X_n) \cup (X_1 \times C(F_2) \times \ldots \times X_n) \\ &\qquad \cup \ldots \cup (X_1 \times X_2 \times \ldots \times C(F_n)) \end{aligned}$$

is a finite union of open sets.

Theorem 8.9 A subset F of X is closed if and only if F is the intersection of a family $(F_\alpha)_{\alpha \in I}$ of sets, where for each $\alpha \in I$, F_α is a finite union of sets of the form

$$F_1 \times F_2 \times \ldots \times F_n,$$

F_i a closed subset of X_i, $i = 1, 2, \ldots, n$.

Proof. We have seen that a set of the form $F_1 \times F_2 \times \ldots \times F_n$, each F_i a closed subset of X_i, $i = 1, 2, \ldots, n$, is a closed subset of X. If, for $\alpha \in I$, F_α is a finite union of such sets, then F_α is closed, and, consequently,

$$\cap_{\alpha \in I} F_\alpha$$

is closed.

Conversely, suppose that F is a closed subset of X, so that $C(F)$ is open. By Theorem 8.5,

$$C(F) = \cup_{\alpha \in I} O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n},$$

where for each $\alpha \in I$, $O_{\alpha,i}$ is an open subset of X_i. Thus,

$$F = C(\cup_{\alpha \in I} O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n})$$

and, by DeMorgan's law,

$$F = \cap_{\alpha \in I} C(O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n}).$$

But, for each $\alpha \in I$,

$$\begin{aligned} &C(O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n}) = (C(O_{\alpha,1}) \times X_2 \times \ldots \times X_n) \\ &\cup (X_1 \times C(O_{\alpha,2}) \times \ldots \times X_n) \cup \ldots \cup (X_1 \times X_2 \times \ldots \times C(O_{\alpha,n})), \end{aligned}$$

so that each of the sets $C(O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n})$ is a finite union of closed sets of the form

$$F_1 \times F_2 \times \ldots \times F_n,$$

F_i closed in X_i, $i = 1, 2, \ldots, n$.

The description of the closed sets of the product space X is more complicated than the description of the open sets. By analogy with the earlier situation, we shall define a "basis" for the closed sets of X to be the sets that play the role of the F_α in the above proof.

Definition 8.10 Let (Y, d') be a metric space and let $\mathfrak{F}$ be a collection of subsets of Y. The collection $\mathfrak{F}$ is called a *basis for the closed sets of* (Y, d') provided that:

1. Each subset in $\mathfrak{F}$ is a closed set;
2. A subset W of Y is closed if and only if W is the intersection of sets belonging to the collection $\mathfrak{F}$.

Theorem 8.9 then becomes:

Corollary 8.11 A basis for the closed sets of (X, d) is the collection of subsets of X that are finite unions of the sets of the form

$$F_1 \times F_2 \times \ldots \times F_n,$$

where for $i = 1, 2, \ldots, n$, F_i is a closed subset of X_i.

The term "sub-base" is used to describe the sets that play the role of $F_1 \times F_2 \times \ldots \times F_n$.

Definition 8.12 Let (Y, d') be a metric space. Let $\mathcal{G}$ be a collection of subsets of Y. The collection $\mathcal{G}$ is called a *sub-base for the closed sets of* (Y, d'), provided that:

1. Each subset in $\mathcal{G}$ is a closed set;
2. A subset W of Y is closed if and only if W is the intersection of sets each of which is a finite union of sets in $\mathcal{G}$, or in place of 2,

2′. The collection $\mathfrak{F}$ of sets that are finite unions of sets belonging to $\mathcal{G}$ is a basis for the closed sets of (Y, d').

Corollary 8.13 A sub-base for the closed sets of (X, d) is the collection of subsets of X of the form

$$F_1 \times F_2 \times \ldots \times F_n,$$

where, for $i = 1, 2, \ldots, n$, F_i is a closed subset of X_i.

Exercises

1. Do Problem 2 in Section 6 again.

2. Do Problems 1 and 2 in Section 7 again.

3. Let F be the subset of (R^2, d) consisting of all the points (x_1, x_2) such that both x_1 and x_2 are integers. Prove that F is closed.

4. Let Q^2 be the subset of (R^2, d) consisting of all points (x_1, x_2) such that both x_1 and x_2 are rational numbers. Prove that Q^2 is neither open nor closed.

5. Let (X, d) be a metric space. Define a distance function d^* on $X \times X$ by the method of Theorem 2.3. Prove that the function $d:(X \times X, d^*) \to (R, d)$ is continuous.

6. In (R^k, d), let $a \in R^k$ and $N_n = S\left(a; \frac{1}{n}\right)$. Prove that $N_1, N_2, \ldots, N_n, \ldots$ is a basis for the neighborhoods at a.

7. In (R^2, d), let A be the subset of R^2 consisting of all those points $x = (x_1, x_2)$ such that $x_2 < x_1$. Sketch the region A and determine a collection $(O_\alpha)_{\alpha \in I}$ of subsets of R^2 of the form $O_{\alpha,1} \times O_{\alpha,2}$, where $O_{\alpha,i}$, $\alpha \in I$, $i = 1, 2$, is open in R, such that $A = \bigcup_{\alpha \in I} O_\alpha$.

9 Subspaces

Let (X, d) be a metric space. Given a subset Y of X we may convert Y into a metric space by restricting the distance function d to $Y \times Y$. In this manner each subset Y of X gives rise to a new metric space $(Y, d \mid Y \times Y)$. On the other hand, we may be given two metric spaces (X, d) and (Y, d'). If $Y \subset X$, it makes sense to ask whether or not d' is the restriction of d.

Definition 9.1 Let (X, d) and (Y, d') be metric spaces. We say that (Y, d') is a *subspace* of (X, d) if:

1. $Y \subset X$;
2. $d' = d \mid Y \times Y$.

Let $Y \subset X$ and $i: Y \to X$ be an inclusion mapping. Denote by $i \times i: Y \times Y \to X \times X$ the inclusion mapping defined by

$$(i \times i)(y_1, y_2) = (y_1, y_2).$$

Then (Y, d') is a subspace of (X, d) if the diagram

$$\begin{tikzcd} Y \times Y \arrow[dr, "d'"] \arrow[d, "i \times i"'] & \\ X \times X \arrow[r, "d"'] & R \end{tikzcd}$$

is commutative. There are as many subspaces of a metric space (X, d) as there are subsets of X.

Example 1 Let Q be the set of rational numbers. Define $d_Q: Q \times Q \to R$ by $d_Q(a, b) = |a - b|$. Then (Q, d_Q) is a subspace of (R, d).

Example 2 Let I^n (the unit n-cube) be the set of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers such that $0 \leqq x_i \leqq 1$, for $i = 1, 2, \ldots, n$. Define $d_c: I^n \times I^n \to R$ by $d_c((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \underset{1 \leqq i \leqq n}{\text{maximum}} \{|x_i - y_i|\}$. Then (I^n, d_c) is a subspace of (R^n, d).

Example 3 Let S^n (the n-sphere) be the set of all $(n + 1)$-tuples $(x_1, x_2, \ldots, x_{n+1})$ of real numbers such that

$$x_1^2 + x_2^2 + \ldots + x_{n+1}^2 = 1.$$

Define $d_S: S^n \times S^n \to R$ by

$$d((x_1, x_2, \ldots, x_{n+1}), (y_1, y_2, \ldots, y_{n+1})) = \sqrt{\sum_{i=1}^{n+1} (x_i - y_i)^2}.$$

Then (S^n, d_S) is a subspace of (R^{n+1}, d').

Example 4 Let A be the set of all $(n + 1)$-tuples $(x_1, x_2, \ldots, x_{n+1})$ of real numbers such that $x_{n+1} = 0$. Define $d_A: A \times A \to R$ by

$$d_A((x_1, x_2, \ldots, x_n, 0), (y_1, y_2, \ldots, y_n, 0)) = \underset{1 \leqq i \leqq n}{\text{maximum}} \{|x_i - y_i|\}.$$

Then (A, d_A) is a subspace of (R^{n+1}, d).

Theorem 9.2 Let (Y, d') be a subspace of (X, d). Then the inclusion mapping $i: Y \to X$ is continuous.

Proof. Given $a \in Y$ and $\varepsilon > 0$, choose $\delta = \varepsilon$. If $d'(a, y) < \delta$, then $d(i(a), i(y)) = d(a, y) = d'(a, y) < \delta = \varepsilon$.

This theorem could also be proved by using the characterization of continuity in terms of neighborhoods, open sets, and so on. To prove this result using neighborhoods we would need to know, for a point $a \in Y$, the complete system of neighborhoods of the point a when a is considered to be a point of (Y, d') and when a is considered to be a point of (X, d). For this purpose we require the relationship between open spheres in (Y, d') and open spheres in (X, d). Let $S_X(a; \delta)$ stand for the set of all points $x \in X$ such that $d(a, x) < \delta$ and $S_Y(a; \delta)$ stand for the set of all points $y \in Y$ such that $d'(a, y) < \delta$. Since d' will be the restriction of d, we obtain:

Lemma 9.3 Let (Y, d') be a subspace of (X, d). Then for a point $a \in Y$,

$$S_Y(a; \delta) = Y \cap S_X(a; \delta).$$

Theorem 9.4 Let (Y, d') be a subspace of (X, d). For a point $a \in Y$, a subset $N' \subset Y$ is a neighborhood of a if and only if there is a neighborhood N of a in (X, d) such that

$$N' = Y \cap N.$$

Proof. First let N be a neighborhood of a in (X, d), so that for some $\delta > 0$, $S_X(a; \delta) \subset N$. Then the subset $N' = Y \cap N$ of Y is a neighborhood of a in (Y, d'), for $S_Y(a; \delta) = Y \cap S_X(a; \delta) \subset N'$.

Conversely, suppose N' is a neighborhood of a in (Y, d'), so that for some $\delta > 0$, $S_Y(a; \delta) \subset N'$. Let $N = N' \cup S_X(a; \delta)$. Then N is a neighborhood of a in (X, d), since $S_X(a; \delta) \subset N$, and

$$\begin{aligned} Y \cap N = Y \cap (N' \cup S_X(a; \delta)) &= (Y \cap N') \cup (Y \cap S_X(a; \delta)) \\ &= N' \cup S_Y(a; \delta) = N'. \end{aligned}$$

The above theorem states that the neighborhoods of a point a in a subspace are precisely the restrictions of the neigh-

borhoods of the point a in the larger space. A similar statement holds for open sets.

Theorem 9.5 Let (Y, d') be a subspace of (X, d). A subset $O' \subset Y$ is an open subset of (Y, d') if and only if there is an open subset O of (X, d) such that

$$O' = Y \cap O.$$

Proof. Let O be an open subset of (X, d) and $O' = Y \cap O$. For each $a \in O'$, since $a \in O$ there is a $\delta_a > 0$ such that $S_X(a; \delta_a) \subset O$. Thus, $S_Y(a; \delta_a) = Y \cap S_X(a; \delta_a) \subset Y \cap O = O'$. Thus O' is open in (Y, d').

Conversely, suppose O' is open in (Y, d'). Then for each $a \in O'$, there is a $\delta_a > 0$ such that $S_Y(a; \delta_a) \subset O'$. The set $S_X(a; \delta_a)$ is open in (X, d) and

$$O = \cup_{a \in O'} S_X(a; \delta_a)$$

is open in (X, d). But

$$Y \cap O = \cup_{a \in O'} (Y \cap S_X(a; \delta_a)) = \cup_{a \in O'} S_Y(a; \delta_a) = O'.$$

Theorem 9.6 Let (Y, d') be a subspace of (X, d). A subset $F' \subset Y$ is a closed subset of (Y, d') if and only if there is a closed subset F of (X, d) such that

$$F' = Y \cap F.$$

Proof. We shall use the characterization of a closed subset F' of Y as a set for which $d'(a, F') = 0$ implies $a \in F'$. First, suppose that $F' = Y \cap F$, where F is a closed subset of X. If $a \in Y$ and $d'(a, F') = 0$, then $d(a, F) = 0$ and $a \in F$. Thus $a \in Y \cap F = F'$. It follows that F' is closed.

Conversely, suppose F' is a closed subset of Y. Let F be the subset of X consisting of those points x such that $d(x, F') = 0$. We claim that F is closed in X and $F' = Y \cap F$. Clearly, $F' \subset F$, so that $F' \subset Y \cap F$. On the other hand, if $x \in Y \cap F$, then $d(x, F') = d'(x, F') = 0$ and the fact that F' is closed in (Y, d') implies that $x \in F'$. Thus $F' = Y \cap F$. Finally, F is closed, for given a point $a \in C(F)$, $d(a, F') > 0$; hence the open sphere of radius $d(a, F')/2$ with center a is contained in $C(F)$, whence $C(F)$ is open.

Exercises

1. Let (Y, d') be a subspace of (X, d). Prove that the inclusion mapping $i: Y \to X$ is continuous by using the characterization of continuity in terms of neighborhoods, in terms of open sets, and in terms of closed sets.
2. Prove that if (Z, d'') is a subspace of (Y, d') and (Y, d') is a subspace of (X, d), then (Z, d'') is a subspace of (X, d).
3. Let (Y, d') be a subspace of (X, d). Let $a_1, a_2, \ldots$ be a sequence of points of Y and let $a \in Y$. Prove that if $\lim_n a_n = a$ in (Y, d'), then $\lim_n a_n = a$ in (X, d). [The converse is false unless one assumes that all the points mentioned lie in Y; see the next problem.]
4. Consider the subspace (Q, d_Q) (the rational numbers) of (R, d). Let $a_1, a_2, \ldots$ be a sequence of rational numbers such that $\lim_n a_n = \sqrt{2}$. Prove that, given $\varepsilon > 0$, there is a positive integer N such that for $n, m > N$, $|a_n - a_m| < \varepsilon$. Does the sequence $a_1, a_2, \ldots$ converge when considered to be a sequence of points of (Q, d_Q)?

10 Equivalence of Metric Spaces

Let A be the set of all $(n + 1)$-tuples $(x_1, x_2, \ldots, x_{n+1})$ of real numbers such that $x_{n+1} = 0$. Thus $A \subset R^{n+1}$. Let d_A be the restriction of the distance function d on R^{n+1} to A, so that (A, d_A) is a subspace of (R^{n+1}, d). The metric space (A, d_A) is in most respects a copy of the metric space (R^n, d). The only distinction between (R^n, d) and (A, d_A) is that a point of R^n is an n-tuple of real numbers, whereas a point of A is an $(n + 1)$-tuple of real numbers of which the last one is zero. The relationship between the metric spaces (R^n, d) and (A, d_A) is an example of the relationship called "metric equivalence."

Definition 10.1 Two metric spaces (A, d_A) and (B, d_B) are said to be *metrically equivalent* if there are inverse functions $f:A \to B$ and $g:B \to A$ such that, for each $x, y \in A$,

$$d_B(f(x), f(y)) = d_A(x, y).$$

and, for each $u, v \in B$,

$$d_A(g(u), g(v)) = d_B(u, v).$$

In this event we shall say that the *metric equivalence is defined by f and g.*

Theorem 10.2 A necessary and sufficient condition that two metric spaces (A, d_A) and (B, d_B) be metrically equivalent is that there exist a function $f:A \to B$ such that:

1. f is one-one;
2. f is onto;
3. for each $x, y \in A$,

$$d_B(f(x), f(y)) = d_A(x, y).$$

Proof. The stated conditions are necessary, for if (A, d_A) and (B, d_B) are metrically equivalent, there are inverse functions $f:A \to B$ and $g:B \to A$, and therefore f is one-one and onto. Conversely, suppose a function $f:A \to B$ with the stated properties exists. Then f is invertible and the function $g:B \to A$ such that f and g are inverse functions is determined by setting $g(b) = a$ if $f(a) = b$. For $u, v \in B$, let $x = g(u)$, $y = g(v)$. Then

$$d_A(g(u), g(v)) = d_A(x, y) = d_B(f(x), f(y)) = d_B(u, v).$$

Given metric spaces (A, d_A) and (B, d_B) and functions $f:A \to B$ and $g:B \to A$, let us denote by $f \times f:A \times A \to B \times B$ the function defined by setting $(f \times f)(x, y) = (f(x), f(y))$ for $x, y \in A$ and, similarly, let $g \times g:B \times B \to A \times A$ be defined by setting $(g \times g)(u, v) = (g(u), g(v))$ for $u, v \in B$. The statement that $d_B(f(x), f(y)) = d_A(x, y)$ for $x, y \in A$ is equivalent to the statement that the diagram

$$\begin{tikzcd}
A \times A \arrow[dr, "d_A"] \arrow[d, "f \times f"'] & \\
B \times B \arrow[r, "d_B"'] & R
\end{tikzcd}$$

is commutative (one may also describe this relation by saying that the function $f:A \to B$ is "distance preserving"). In terms of diagrams, the statement that (A, d_A) and (B, d_B) are metrically equivalent is the statement that there exist functions $f:A \to B$, $g:B \to A$ such that the four diagrams

$$\begin{tikzcd}
A \arrow[rr, "i_A"] \arrow[dr, "f"'] & & A \\
& B \arrow[ur, "g"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
B \arrow[rr, "i_B"] \arrow[dr, "g"'] & & B \\
& A \arrow[ur, "f"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
A \times A \arrow[dr, "d_A"] \arrow[d, "f \times f"'] & \\
B \times B \arrow[r, "d_B"'] & R
\end{tikzcd}
\qquad
\begin{tikzcd}
B \times B \arrow[dr, "d_B"] \arrow[d, "g \times g"'] & \\
A \times A \arrow[r, "d_A"'] & R
\end{tikzcd}$$

are commutative (where $i_A:A \to A$ and $i_B:B \to B$ are identity mappings.) The first two diagrams express the fact that f and g are inverse functions and the last two diagrams express the fact that f and g "preserve distances." Since the distance between x and y in A is the same as the distance between $f(x)$ and $f(y)$ in B, f is continuous. Similarly, g is continuous. Thus:

Lemma 10.3 Let a metric equivalence between (A, d_A) and (B, d_B) be defined by inverse functions $f:A \to B$ and $g:B \to A$. Then both f and g are continuous.

From the point of view of considerations that relate only to the concept of continuity, the relationship of metric equivalence is too narrow. We are led to define a broader concept of equivalence in which we drop the requirement of "preservation of distance"; that is, the commutativity of the last pair of diagrams, and merely require that the first two diagrams be commutative and the functions in these diagrams be continuous.

Definition 10.4 Two metric spaces (A, d_A) and (B, d_B) are said to be *topologically equivalent* if there are inverse functions $f:A \to B$ and $g:B \to A$ such that f and g are continuous. In this event we say that the *topological equivalence is defined by f and g.*

As a corollary to Lemma 10.3 we obtain:

Corollary 10.5 Two metric spaces that are metrically equivalent are topologically equivalent.

The converse of this corollary is false; that is, there are metric spaces that are topologically equivalent, but are not metrically equivalent. For example, a circle of radius 1 is topologically equivalent to a circle of radius 2 (considered as subspaces of (R^2, d)), but the two are not metrically equivalent.

The following two lemmas furnish a sufficient condition for the topological equivalence of two metric spaces with the same underlying sets.

Lemma 10.6 Let (X, d_1) and (X, d_2) be two metric spaces. If there exists a number $K > 0$ such that for each $x, y \in X$,

$$d_2(x, y) \leqq Kd_1(x, y),$$

then the identity mapping

$$i:(X, d_1) \to (X, d_2)$$

is continuous.

Proof. Given $\varepsilon > 0$ and $a \in X$, set $\delta = \varepsilon/K$. If $d_1(x, a) < \delta$ then $d_2(i(x), i(a)) = d_2(x, a) \leqq K\cdot d_1(x, a) < K\delta = \varepsilon$.

Lemma 10.7 Let (X, d) and (X, d') be two metric spaces. If there exist positive numbers K and K' such that for each $x, y \in X$,

$$d'(x, y) \leqq K\cdot d(x, y),$$

$$d(x, y) \leqq K'\cdot d'(x, y),$$

then the identity mappings define a topological equivalence between (X, d) and (X, d').

We have discussed the two metric spaces (R^n, d) and (R^n, d'), where the distance function d is determined by the maximum distance between coordinates, and the distance function d' is what is called the Euclidean distance function. For each pair of points $x, y \in R^n$, the inequality

$$d(x, y) \leqq d'(x, y) \leqq \sqrt{n}\, d(x, y)$$

holds. It therefore follows from Lemma 10.7 that the metric spaces (R^n, d) and (R^n, d') are topologically equivalent.

Theorem 10.8 Let (X, d) and (Y, d') be two metric spaces. Let $f:X \to Y$ and $g:Y \to X$ be inverse functions. Then the following four statements are equivalent:

1. f and g are continuous;
2. A subset O of X is open if and only if $f(O)$ is an open subset of Y;
3. A subset F of X is closed if and only if $f(F)$ is a closed subset of Y;
4. For each $a \in X$ and subset N of X, N is a neighborhod of a if and only if $f(N)$ is a neighborhood of $f(a)$.

Proof. $1 \Rightarrow 2$. Let O be an open subset of X. Then $f(O) = g^{-1}(O)$ is open since g is continuous. Conversely, if $f(O)$ is an open subset of Y, then $f^{-1}(f(O)) = O$ is open since f is continuous.

$2 \Rightarrow 4$. For each $a \in X$ and $N \subset X$, N is a neighborhood of a if and only if N contains an open set O containing a if and only if

$f(N)$ contains an open set $O' = f(O)$ containing $f(a)$ if and only if $f(N)$ is a neighborhood of $f(a)$.

$4 \Rightarrow 1$. For each $a \in X$, let U be a neighborhood of $f(a)$. Then $f^{-1}(U)$ is a neighborhood of a, for $U = f(f^{-1}(U))$ is a neighborhood of $f(a)$. Thus f is continuous. Similarly, for each $b \in Y$, let V be a neighborhood of $g(b)$. Then $g^{-1}(V) = f(V)$ is a neighborhood of $f(g(b)) = b$, and g is continuous.

Thus, statements 1, 2, and 4 are equivalent. We leave it to the reader to verify that statements 2 and 3 are equivalent.

Statement 1 in Theorem 10.8 is, of course, the statement that the metric spaces (X, d) and (Y, d') are topologically equivalent. Consequently, Theorem 10.8 asserts that two metric spaces are topologically equivalent if and only if there exist inverse functions that establish either a one-one correspondence between the open sets of the two spaces, a one-one correspondence between the closed sets of the two spaces, or a one-one correspondence between the complete systems of neighborhoods of the two spaces.

Exercises

1. Prove that each metric space (X, d) is topologically equivalent to itself.
2. Prove that if (X, d) and (Y, d') are topologically equivalent, then any metric space topologically equivalent to (X, d) is also topologically equivalent to (Y, d').
3. For each pair of points $a, b \in R^n$, prove that there is a topological equivalence between (R^n, d) and itself defined by inverse functions $f: R^n \to R^n$ and $g: R^n \to R^n$ such that $f(a) = b$. [*Hint:* If $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots, b_n)$, define f by setting $f(x_1, x_2, \ldots, x_n) = (x_1 + b_1 - a_1, x_2 + b_2 - a_2, \ldots, x_n + b_n - a_n)$.]
4. Prove that the open interval $(-\pi/2, \pi/2)$, considered as a subspace of the real number system, is topologically equivalent to the real number system. Prove that any two open intervals, considered as subspaces of the real number system, are topologically equivalent. Prove that any open interval, considered as a subspace of the real number system, is topologically equivalent to the real number system.
5. For $i = 1, 2, \ldots, n$, let the metric space (X_i, d_i) be topologically equivalent to the metric space (Y_i, d_i'). Prove that if

$$X = \prod_{i=1}^{n} X_i \quad \text{and} \quad Y = \prod_{i=1}^{n} Y_i$$

are converted into metric spaces in the standard manner, then these two metric spaces are topologically equivalent.

6. The open n-cube is the set of all points $x = (x_1, x_2, \ldots, x_n) \in R^n$ such that $0 < x_i < 1$ for $i = 1, 2, \ldots, n$. Prove that the open n-cube, considered as a subspace of (R^n, d), is topologically equivalent to (R^n, d). [*Hint:* Use the results of Problems 4 and 5.]

Appendix

Theorem 2.5 (R^n, d') is a metric space, where d' is the function defined by the correspondence

$$d'(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

for $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n) \in R^n$.

Proof. It is readily apparent that d' possesses the properties required of a distance function, with the exception of the property

$$d'(a, b) \leq d'(a, c) + d'(c, b)$$

for $a, b, c \in R^n$. Let $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots, b_n)$, $c = (c_1, c_2, \ldots, c_n)$. We must show that

$$\sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \leq \sqrt{\sum_{i=1}^{n} (a_i - c_i)^2} + \sqrt{\sum_{i=1}^{n} (c_i - b_i)^2}.$$

If we set $u_i = a_i - c_i$ and $v_i = c_i - b_i$, then $a_i - b_i = u_i + v_i$, and we must show that

$$\sqrt{\sum_{i=1}^{n} (u_i + v_i)^2} \leq \sqrt{\sum_{i=1}^{n} u_i^2} + \sqrt{\sum_{i=1}^{n} v_i^2}.$$

Since the terms on both sides of this inequality are positive, it suffices to prove that

$$\sum_{i=1}^{n} (u_i + v_i)^2 \leqq \sum_{i=1}^{n} u_i^2 + \sqrt{\sum_{i=1}^{n} u_i^2}\sqrt{\sum_{i=1}^{n} v_i^2} + \sum_{i=1}^{n} v_i^2.$$

This last inequality is equivalent to the inequality

$$\sum_{i=1}^{n} u_i v_i \leqq \left[\sum_{i=1}^{n} u_i^2\right]^{1/2} \left[\sum_{i=1}^{n} v_i^2\right]^{1/2}.$$

The proof of this theorem will have been completed when we prove the following lemma:

Schwarz's Lemma Let $(u_1, u_2, \ldots, u_n)$, $(v_1, v_2, \ldots, v_n)$ be n-tuples of real numbers, then

$$\sum_{i=1}^{n} u_i v_i \leqq \left[\sum_{i=1}^{n} u_i^2\right]^{1/2} \left[\sum_{i=1}^{n} v_i^2\right]^{1/2}.$$

Proof. It suffices to prove that

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leqq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right).$$

To this end, we consider, for an arbitrary real number λ, the expression $\sum_{i=1}^{n} (u_i + \lambda v_i)^2$. We have,

$$0 \leqq \sum_{i=1}^{n} (u_i + \lambda v_i)^2 = \sum_{i=1}^{n} u_i^2 + 2\lambda \sum_{i=1}^{n} u_i v_i + \lambda^2 \sum_{i=1}^{n} v_i^2.$$

Therefore, the quadratic equation in λ,

$$0 = \sum_{i=1}^{n} u_i^2 + 2\lambda \sum_{i=1}^{n} u_i v_i + \lambda^2 \sum_{i=1}^{n} v_i^2,$$

can have at most one real solution. Consequently,

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 - \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right) \leqq 0,$$

or

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leqq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right).$$

Topological Spaces

III

1 Introduction

In the context of metric spaces, the various topological concepts such as continuity, neighborhood, and so on, may be characterized by means of open sets. Discarding the distance function and retaining the open sets of a metric space gives rise to a new mathematical object, called a *topological space*. The topological concepts that have been studied in Chapter II must be reintroduced in the context of topological spaces. The procedure for formulating the appropriate definitions of these terms in a topological space is to find, in a metric space, the characterization of the term by means of open sets, using in most cases what is a theorem in a metric space as a definition in a topological space. There are other ways of introducing topological spaces. For example, if, upon discarding the distance function of a metric space, we were to retain the

systems of neighborhoods of the points of the metric space, we obtain what we shall call a neighborhood space. We shall indicate the equivalence between the concept of a neighborhood space and the concept of a topological space. Certain new topological concepts are also introduced; namely, the closure, interior, and boundary of a set (these concepts could have been introduced in metric spaces.) In many respects the elementary material in this chapter is a repetition of material from Chapter II, but in a different context. The concept of a topological space is one of the most fruitful concepts of modern mathematics. It is the proper setting for discussions based on considerations of continuity.

2 Topological Spaces

Definition 2.1 Let X be a non-empty set and $\mathfrak{I}$ a collection of subsets of X such that:

O1. $X \in \mathfrak{I}$.

O2. $\emptyset \in \mathfrak{I}$.

O3. If $O_1, O_2, \ldots, O_n \in \mathfrak{I}$, then

$$O_1 \cap O_2 \cap \cdots \cap O_n \in \mathfrak{I}.$$

O4. If for each $\alpha \in I$, $O_\alpha \in \mathfrak{I}$, then

$$\bigcup_{\alpha \in I} O_\alpha \in \mathfrak{I}.$$

The pair of objects $(X, \mathfrak{I})$ is called a *topological space*. The set X is called the *underlying set* and the collection $\mathfrak{I}$ is called the *topology* on the set X.

Let (X, d) be a metric space. The collection $\mathfrak{I}$ of open sets of this metric space satisfies the conditions *O1*, *O2*, *O3*, *O4* (by Theorem 5.5, Chapter II). Thus, (X, d) gives rise to the topological space $(X, \mathfrak{I})$.

Definition 2.2 Let (X, d) be a metric space. Let $\mathfrak{I}$ be the collection of open sets of this metric space. The topological space $(X, \mathfrak{I})$ is called the *topological space associated with the metric space* (X, d) and the metric space (X, d) is said to *give rise to the topological space* $(X, \mathfrak{I})$.

Thus, we are in a position to give many examples of topological spaces; namely, for each metric space its associated topological space. On the other hand, any set X and collection $\mathfrak{I}$ of subsets satisfying *O1*, *O2*, *O3*, *O4* is an example of a topological space, and we shall see that not every such example arises from a metric space.

Examples

1. The *real line*, that is, the topological space that arises from the metric space consisting of the real number system and the distance function $d(a, b) = |a - b|$.
2. The topological space that arises from the metric space (R^n, d). We shall call this topological space *Euclidean n-space with the usual topology*.
3. Let X be an arbitrary set. Let $\mathfrak{I} = \{\emptyset, X\}$. Then $(X, \mathfrak{I})$ is a topological space.
4. Let X be a set containing precisely two distinct elements a and b. Let $\mathfrak{I}_1 = \{\emptyset, X\}$, $\mathfrak{I}_2 = \{\emptyset, \{a\}, X\}$, $\mathfrak{I}_3 = \{\emptyset, \{b\}, X\}$, $\mathfrak{I}_4 = \{\emptyset, \{a\}, \{b\}, X\}$. Then $(X, \mathfrak{I}_i)$, $i = 1, 2, 3, 4$ are four distinct topological spaces with the same underlying set.
5. Let X be an arbitrary set. Let $\mathfrak{I}$ be the collection of all subsets of X, i.e., $\mathfrak{I} = 2^X$. Then $(X, \mathfrak{I})$ is a topological space. Of all the various topologies that one may place on a set X, this one contains the largest number of elements and is called the *discrete* topology.
6. Let X be an arbitrary set. Let $\mathfrak{I}$ be the collection of all subsets of X whose complements are either finite or all of X. Then $(X, \mathfrak{I})$ is a topological space.
7. Let Z be the set of positive integers. For each positive integer n, let $O_n = \{n, n + 1, n + 2, \ldots\}$. Let $\mathfrak{I} = \{\emptyset, O_1, O_2, \ldots, O_n, \ldots\}$. Then $(Z, \mathfrak{I})$ is a topological space.

To verify that $(X, \mathfrak{J})$ is a topological space, one verifies that the specified collection of subsets, $\mathfrak{J}$, is a topology; that is, that $\mathfrak{J}$ satisfies conditions $O1$, $O2$, $O3$, $O4$. For example, let X and $\mathfrak{J}$ be as in Example 6. Then $X \in \mathfrak{J}$, for its complement $\emptyset = C(X)$ is certainly finite. Also $\emptyset \in \mathfrak{J}$, since $C(\emptyset) = X$. Thus, $\mathfrak{J}$ satisfies conditions $O1$ and $O2$. Next, let $O_1, O_2, \ldots, O_n$ be subsets of X, each of whose complements is finite or all of X. To show that $O_1 \cap O_2 \cap \ldots \cap O_n \in \mathfrak{J}$ we must show that $C(O_1 \cap O_2 \cap \ldots \cap O_n)$ is either finite or all of X. But $C(O_1 \cap O_2 \cap \ldots \cap O_n) = C(O_1) \cup C(O_2) \cup \ldots \cup C(O_n)$. Either this set is a union of finite sets and hence finite, or for some i, $C(O_i) = X$ and the union is all of X. Finally, for each $\alpha \in I$, let $O_\alpha \in \mathfrak{J}$, so that $C(O_\alpha)$ is either finite or X. Then $C(\cup_{\alpha \in I} O_\alpha) = \cap_{\alpha \in I} C(O_\alpha)$. Either each of the sets, $C(O_\alpha) = X$, in which case the intersection is all of X, or at least one of them is finite, in which case the intersection is a subset of a finite set and hence finite. Thus $(X, \mathfrak{J})$ is a topological space. The reader should verify that the remaining examples do, in fact, constitute examples of topological spaces.

Given a topological space $(X, \mathfrak{J})$, the subsets O of X that belong to $\mathfrak{J}$ are called "open" sets. The adjective "open" is used because, in the event that the topological space $(X, \mathfrak{J})$ arises from a metric space (X, d), the subsets of X that are open in the associated topological space $(X, \mathfrak{J})$ are precisely those subsets of X which are open in the metric space (X, d). One may describe this situation by picturing the totality of metric spaces being related to a subcollection of the totality of topological spaces, each metric space (X, d) giving rise to its associated topological space $(X, \mathfrak{J})$, as indicated in Figure 11. We shall see that two distinct metric spaces (X, d) and (X, d') may give rise to the same topological space $(X, \mathfrak{J})$. Also there are topological spaces $(Y, \mathfrak{J}')$, such as Example 7 above, which could not have arisen from a metric space. The subcollection of topological spaces that arise from metric spaces is called the collection of *metrizable* topological spaces.

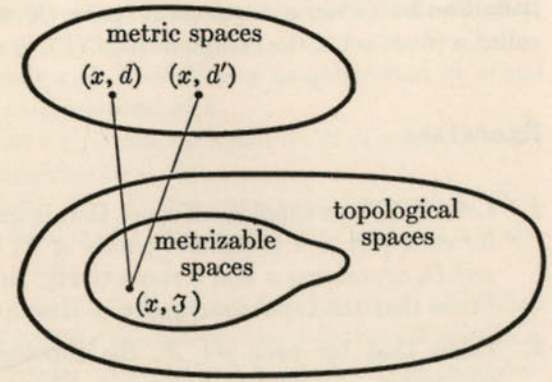


Figure 11

In passing from a metric space to its associated topological space, we may say that the "open" sets have been "preserved."

Definition 2.3 Given a topological space $(X, \mathfrak{J})$, a subset N of X is called a *neighborhood* of a point $a \in X$ if N contains an open set that contains a.

This definition has been formulated so that a subset N of a metric space (X, d) is a neighborhood of a point $a \in X$ if and only if N is a neighborhood of a in the associated topological space. Thus, in passing from a metric space to a topological space, neighborhoods have also been "preserved."

Corollary 2.4 Let $(X, \mathfrak{J})$ be a topological space. A subset O of X is open if and only if O is a neighborhood of each of its points.

Proof. First, suppose that O is open. Then, for each $x \in O$, O contains an open set containing x; namely, O itself. Conversely, suppose O is a neighborhood of each of its points. Then for each $x \in O$, there is an open set O_x such that $x \in O_x \subset O$. Consequently,

$$O = \cup_{x \in O} O_x$$

is a union of open sets and hence open.

Definition 2.5 Given a topological space $(X, \mathfrak{I})$, a subset F of X is called a *closed* set if the complement, $C(F)$, is an open set.

Exercises

1. Let $(X, \mathfrak{I})$ be a topological space that is metrizable. Prove that for each pair a, b of distinct points of X, there are open sets O_a and O_b containing a and b respectively, such that $O_a \cap O_b = \emptyset$. Prove that the topological space of Example 7 is not metrizable.
2. Prove that for each set X, the topological space $(X, 2^X)$ is metrizable. [*Hint:* See Exercise 2, Chapter II, Section 5.]
3. Prove that the two metric spaces (R^n, d) and (R^n, d') give rise to the same topological space.
4. Let $(X, \mathfrak{I})$ be a topological space. Prove that $\emptyset$, X are closed sets, that a finite union of closed sets is a closed set, and that an arbitrary intersection of closed sets is a closed set.
5. Let $(X, \mathfrak{I})$ be a topological space that is metrizable. Prove that each neighborhood N of a point $a \in X$ contains a neighborhood V of a such that V is a closed set.
6. Prove that in a discrete topological space, each subset is simultaneously open and closed.

3 Neighborhoods and Neighborhood Spaces

Theorem 4.9, Chapter II, in which are stated certain properties of neighborhoods in a metric space, corresponds to a theorem in topological spaces.

Theorem 3.1 Let $(X, \mathfrak{I})$ be a topological space.

$N0$. For each point $x \in X$, there is at least one neighborhood N of x.

$N1$. For each point $x \in X$ and each neighborhood N of x, $x \in N$.

$N2$. For each point $x \in X$, if N is a neighborhood of x and $N' \supset N$, then N' is a neighborhood of x.

$N3$. For each point $x \in X$ and each pair N, M of neighborhoods of x, $N \cap M$ is also a neighborhood of x.

$N4$. For each point $x \in X$ and each neighborhood N of x, there exists a neighborhood O of x such that $O \subset N$ and O is a neighborhood of each of its points.

Proof. For each point $x \in X$, X is a neighborhood of x, thus $N0$ is true. $N1$ and $N2$ follow easily from the definition of neighborhood in a topological space. To verify $N3$, let N, M be neighborhoods of x. Then there are open sets O and O' such that $N \supset O$, $M \supset O'$ and $x \in O$, $x \in O'$. Thus, $N \cap M$ contains the open set $O \cap O'$, which contains x, and, consequently, $N \cap M$ is a neighborhood of x. Finally, for a point $x \in X$, let N be a neighborhood of x. Then N contains an open set O containing x. In particular, O is a neighborhood of x and by Corollary 2.4, O is a neighborhood of each of its points.

In a topological space, as in a metric space, we lay down the definition:

Definition 3.2 For each point x in a topological space $(X, \mathfrak{I})$, the collection $\mathfrak{N}_x$ of all neighborhoods of x is called a *complete system of neighborhoods at the point* x.

One may paraphrase the properties $N0$–$N4$ of neighborhoods in terms of the complete system of neighborhoods $\mathfrak{N}_x$ at the points $x \in X$:

$N0$. For each $x \in X$, $\mathfrak{N}_x \neq \emptyset$;

$N1$. For each $x \in X$ and $N \in \mathfrak{N}_x$, $x \in N$;

$N2$. For each $x \in X$ and $N \in \mathfrak{N}_x$, if $N' \supset N$ then $N' \in \mathfrak{N}_x$;

$N3$. For each $x \in X$ and $N, M \in \mathfrak{N}_x$, $N \cap M \in \mathfrak{N}_x$;

$N4$. For each $x \in X$ and $N \in \mathfrak{N}_x$, there exists an $O \in \mathfrak{N}_x$ such that $O \subset N$ and $O \in \mathfrak{N}_y$ for each $y \in O$.

The proof of Theorem 3.1 was, in most respects, similar to the proof of the corresponding theorem in metric spaces,

Theorem 4.9, Chapter II. However, it was necessary to supply a proof of Theorem 3.1 above, for in the proof of 4.9, Chapter II, use was made of the concept of open spheres, a concept which does not occur in a topological space. Though a comparison of these two theorems might lead one to believe that statements about neighborhoods that are true in a metric space are also true in a topological space, this is not always the case. We have seen (Theorem 4.10, Chapter II) that in a metric space (X, d), given two distinct points x, y, there are neighborhoods N and M of x and y respectively, such that $N \cap M = \emptyset$. This statement is false in many topological spaces. For example, let $Y = \{a, b\}$, $a \neq b$, and let $\mathfrak{I} = \{\emptyset, \{a\}, Y\}$, so that $(Y, \mathfrak{I})$ is a topological space. Then the only neighborhood of b is Y. Thus, for each neighborhood N of a and each neighborhood M of b, $N \cap M = N \cap Y = N \neq \emptyset$.

Definition 3.3 A topological space $(X, \mathfrak{I})$ is called a *Hausdorff space* or is said to satisfy the *Hausdorff axiom*, if for each pair a, b of distinct points of X, there are neighborhoods N and M of a and b respectively, such that $N \cap M = \emptyset$.

Some authors use the term "separated space" instead of Hausdorff space. Most of the significant topological spaces are Hausdorff spaces. For this reason certain authors require a topological space to be a Hausdorff space and use the two terms synonymously; that is, they add to the list $O1$–$O4$ of properties of open sets in the definition of a topological space, the property, for each pair x, y of distinct points there are open sets O_x and O_y containing x and y respectively, such that $O_x \cap O_y = \emptyset$.

Suppose we have a metric space (X, d) and we discard the distance function, retaining only the neighborhoods of the points in X. Then for each point $x \in X$, we have a collection $\mathfrak{N}_x$ of subsets of X; namely the complete system of neighborhoods at x. These neighborhoods satisfy certain properties. We may select some of these properties and use them as a set of axioms for what we might naturally call a "neighborhood space."

Definition 3.4 Let X be a set. For each $x \in X$, let there be given a collection $\mathfrak{N}_x$ of subsets of x (called the neighborhoods of x), satisfying the conditions $N0$–$N4$ of Theorem 3.2. This object is called a *neighborhood space*.

In a neighborhood space, the appropriate definition of open set is obtained from Corollary 2.4.

Definition 3.5 In a neighborhood space, a subset O is said to be open if it is a neighborhood of each of its points.

It is important to realize that the mathematical object neighborhood space, although closely connected with the concept of a topological space, is a new object, and until we have defined the term *open set* in a neighborhood space, that term in a neighborhood space is meaningless.

Lemma 3.6 In a neighborhood space, the empty set and the whole space are open, a finite intersection of open sets is open, and an arbitrary union of open sets is open.

Proof. [Since we are concerned with neighborhood spaces, we may use only the properties $N0$–$N4$ of neighborhoods and, of course, Definition 3.5 of open sets.] The empty set is open, for in order for it not to be open it would have to contain a point x of which it was not a neighborhood. Given a point x, there is some neighborhood N of x, so by $N2$, the whole space is a neighborhood of x. Thus, the whole space is a neighborhood of each of its points and hence open. If O and O' are open, then $O \cap O'$ is also open, for by $N3$, given $x \in O \cap O'$, O and O' are neighborhoods of x, hence so is $O \cap O'$. Thus the intersection of two open sets is a neighborhood of each of its points, and, consequently, by induction, any finite intersection of open sets is open. Finally, suppose for each $\alpha \in I$, O_α is open. If

$x \in \bigcup_{\alpha \in I} O_\alpha$, then $x \in O_\beta$ for some $\beta \in I$. But O_β is a neighborhood of x and $O_\beta \subset \bigcup_{\alpha \in I} O_\alpha$, thus by $\mathcal{N}2$, $\bigcup_{\alpha \in I} O_\alpha$ is a neighborhood of x and is therefore open.

If we start with a topological space and define neighborhoods by Definition 2.3, Theorem 3.1 tells us that the underlying set and the complete systems of neighborhoods of the points of the set yield a neighborhood space. On the other hand, if we start with a neighborhood space and define open sets by Definition 3.5, Lemma 3.6 tells us that we obtain a topological space. Suppose then, we have a topological space $(X, \mathfrak{I})$, use the neighborhoods of $(X, \mathfrak{I})$ to form a neighborhood space, and finally use the open sets in this neighborhood space to create a topological space $(X, \mathfrak{I}')$. Do we end up with our original topological space $(X, \mathfrak{I})$? The answer is yes. To prove this result we must show that $\mathfrak{I} = \mathfrak{I}'$. Now, if O is an open set in our original topological space, that is, $O \in \mathfrak{I}$, by Corollary 2.4, O is a neighborhood of each of its points, from which it follows that O is an open subset of the neighborhood space and hence $O \in \mathfrak{I}'$. Conversely, if $O \in \mathfrak{I}'$, then in the neighborhood space, O is a neighborhood of each of its points. But the neighborhoods of the neighborhood space we have created are the neighborhoods of $(X, \mathfrak{I})$, so that again by Corollary 2.4, O is open in $(X, \mathfrak{I})$ or $O \in \mathfrak{I}$. Thus $\mathfrak{I} = \mathfrak{I}'$.

Logically, it would still be possible for there to be neighborhood spaces that did not arise in this manner from topological spaces. We shall now show that there are none. To do so, we need a characterization, in a neighborhood space, of neighborhoods in terms of open sets.

Lemma 3.7 In a neighborhood space, a subset N is a neighborhood of a point x if and only if N contains an open set containing x.

Proof. First, let N contain an open set O containing x. By Definition 3.5, O is a neighborhood of x, whence, by $\mathcal{N}2$, N is a neighborhood of x. Conversely, if N is a neighborhood of x, then by $\mathcal{N}4$, N contains a neighborhood O of x (and by $\mathcal{N}1$, O contains x), such that O is a neighborhood of each of its points.

To denote a neighborhood space, let us use the symbol $(X, \mathfrak{N})$, where for each $x \in X$, $\mathfrak{N}_x$ is the collection of neighborhoods of x. Now suppose that we start with a neighborhood space $(X, \mathfrak{N})$. We define open set in $(X, \mathfrak{N})$ by Definition 3.5, thus obtaining a topological space $(X, \mathfrak{I})$. In the topological space $(X, \mathfrak{I})$ we define neighborhood by Definition 2.3 to obtain a neighborhood space $(X, \mathfrak{N}')$. Under these circumstances, if $N \in \mathfrak{N}_x$, by Lemma 3.7, N contains an open set O containing x, so that by Definition 2.3, N is a neighborhood of x in $(X, \mathfrak{I})$, or $N \in \mathfrak{N}'_x$. Conversely, if $N \in \mathfrak{N}'_x$, then by Definition 2.3, N contains a set $O \in \mathfrak{I}$, and $x \in O$. Since $O \in \mathfrak{I}$, O is open in the neighborhood space $(X, \mathfrak{N})$ and so by Lemma 3.7, N is a neighborhood of x. Thus, for each $x \in X$, $\mathfrak{N}_x = \mathfrak{N}'_x$, and the two neighborhood spaces are the same.

Collecting together the results on the correspondence between topological spaces and neighborhood spaces, we have:

Theorem 3.8 Let neighborhood in a topological space be defined by Definition 2.3 and open set in a neighborhood space be defined by Definition 3.5. Then the neighborhoods of a topological space $(X, \mathfrak{I})$ give rise to a neighborhood space $(X, \mathfrak{N}) = \mathfrak{A}(X, \mathfrak{I})$ and the open sets of a neighborhood space $(Y, \mathfrak{N}')$ give rise to a topological space $(Y, \mathfrak{I}') = \mathfrak{A}'(Y, \mathfrak{N}')$. Furthermore, for each topological space $(X, \mathfrak{I})$,

$$(X, \mathfrak{I}) = \mathfrak{A}'(\mathfrak{A}(X, \mathfrak{I})),$$

and for each neighborhood space $(X, \mathfrak{N})$,

$$(X, \mathfrak{N}) = \mathfrak{A}(\mathfrak{A}'(X, \mathfrak{N})),$$

thus establishing a one-one correspondence between the collection of all topological spaces and the collection of all neighborhood spaces.

Theorem 3.8 justifies the specification of a topological space by defining for a given set X what subsets of X are to be the neighborhoods of a point $x \in X$; that is, by specifying

the corresponding neighborhood space. For example, let X be the set of positive integers. Given a point $n \in X$ and a subset U of X, let us call U a neighborhood of n if for each integer $m \geqq n$, $m \in U$. We must then verify that these neighborhoods satisfy conditions $N0$–$N4$ so that we have a neighborhood space and consequently a topological space. The reader should verify that this corresponding topological space is the one described in Example 7 of Section 2.

Exercises

1. Define a co-topological space to be a set X and a collection $\mathfrak{F}$ of subsets of X such that $\emptyset$ and X are in $\mathfrak{F}$, a finite union of sets in $\mathfrak{F}$ is again in $\mathfrak{F}$, and an arbitrary intersection of sets in $\mathfrak{F}$ is also in $\mathfrak{F}$. Call the elements of $\mathfrak{F}$ closed sets. Prove a theorem, similar to Theorem 3.8, about the relationship between topological spaces and co-topological spaces.

2. Given a real number x, call a subset N of R a neighborhood of x if $y \geqq x$ implies $y \in N$. Prove that this definition of neighborhood yields a neighborhood space. Describe the corresponding topological space.

3. Given a real number x, call a subset N or R a neighborhood of x if N contains the closed interval $[x, x + 1]$. Prove that the neighborhoods so defined satisfy $N0$–$N3$, but not $N4$. Use the Definition 3.5 of open set anyway, and determine which subsets of R will be open.

4. In a neighborhood space, a collection $\mathcal{B}_x$ of neighborhoods of a point $x \in X$ is called a *basis for the complete system of neighborhoods at* x, or simply a *basis for the neighborhoods at* x, if, for each neighborhood N of x, there is a neighborhood $U \in \mathcal{B}_x$ such that

$$U \subset N.$$

Prove that for each point $x \in X$, if $\mathcal{B}_x$ is the collection of open sets containing x, then $\mathcal{B}_x$ is a basis for the neighborhoods at x.

Prove that if for each point $x \in X$, $\mathcal{B}_x$ is a basis for the neighborhoods at x, then:

$BN0$. For each $x \in X$, $\mathcal{B}_x \neq \emptyset$;

$BN1$. For each $x \in X$ and $U \in \mathcal{B}_x$, $x \in U$;

$BN2$. For each $x \in X$ and $U, V \in \mathcal{B}_x$, $U \cap V$ contains an element $W \in \mathcal{B}_x$;

$BN3$. For each $x \in X$ and $U \in \mathcal{B}_x$, there is an $O \subset U$ such that $x \in O$ and for each $y \in O$, O contains an element $V_y \in \mathcal{B}_y$.

Define a *basic neighborhood space* to be a set X and for each $x \in X$ a collection $\mathcal{B}_x$ of subsets of X satisfying $BN0$–$BN3$. In a basic neighborhood space $(X, \mathcal{B})$ define a subset N of X to be a neighborhood of a point $x \in X$, if $N \supset U$ for some $U \in \mathcal{B}_x$. Prove that the neighborhoods of a basic neighborhood space yield a neighborhood space. (Thus a topological space may be constructed by specifying for each point x a basis $\mathcal{B}_x$ of the neighborhoods at x satisfying $BN0$–$BN3$.) The correspondence between basic neighborhood spaces and neighborhood spaces is many-one, since there are many different bases for the neighborhoods at a point in a neighborhood space. However, prove that if $(X, \mathcal{B})$ and $(X, \mathcal{B}')$ are two basic neighborhood spaces, then they give rise to the same neighborhood space if and only if for each point $x \in X$ we have

(i) given $U \in \mathcal{B}_x$, there is a $U' \in \mathcal{B}'_x$ with $U' \subset U$, and
(ii) given $V' \in \mathcal{B}'_x$, there is a $V \in \mathcal{B}_x$ with $V \subset V'$.

Also prove that starting from a given neighborhood space $(X, \mathfrak{N})$, if for each $x \in X$, $\mathcal{B}_x$ is a basis for the neighborhoods at x, then the neighborhood space that arises from the basic neighborhood space $(X, \mathcal{B})$ is $(X, \mathfrak{N})$.

4 Closure, Interior, Boundary

In a metric space, given a point x and a subset A, we can say that there are points of A arbitrarily close to x if $d(x, A) = 0$. In a topological space, we can also find a characterization of "arbitrary closeness." To indicate the proper translation from metric spaces to topological spaces of this concept, let us first prove:

Lemma 4.1 In a metric space (X, d), for a given point x and a given subset A, $d(x, A) = 0$ if and only if each neighborhood N of x contains a point of A.

Proof. First, suppose that each neighborhood N of x contains a point of A. In particular, for each $\varepsilon > 0$, there is a point of A in $S(x; \varepsilon)$. Thus $\underset{a \in A}{\text{l.u.b.}} \{d(x, a)\} < \varepsilon$ for each $\varepsilon > 0$ and consequently $d(x, A) = \underset{a \in A}{\text{l.u.b.}} \{d(x, a)\} = 0$. Conversely, suppose that there is a neighborhood N of x that does not contain a point of A. Since N is a neighborhood of x in a metric space, there is an $\varepsilon > 0$ such that $S(x; \varepsilon) \subset N$. It follows that $a \in A$ implies that $d(x, a) \geqq \varepsilon$. Thus $d(x, A) \geqq \varepsilon$.

We shall, therefore, in a topological space, say that the points of a subset A are arbitrarily close to a given point x, if each neighborhood of x contains a point of A. Given a subset A, the collection of points that are arbitrarily close to A is called the *closure* of A.

Definition 4.2 Let A be a subset of a topological space. A point x is said to be *in the closure of* A if, for each neighborhood N of x, $N \cap A \neq \emptyset$. The closure of A is denoted by $\overline{A}$.

The purpose of the next two lemmas is to provide a description of the closure of a subset in terms of closed sets.

Lemma 4.3 Given a subset A of a topological space and a closed set F containing A,

$$\overline{A} \subset F.$$

Proof. Suppose $x \notin F$, then x is in the open set $C(F)$. Also, $F \supset A$ implies $C(F) \subset C(A)$. Thus, $C(F) \cap A = \emptyset$. Since $C(F)$ is a neighborhood of x, $x \notin \overline{A}$. We have thus shown that $C(F) \subset C(\overline{A})$ or $\overline{A} \subset F$.

Lemma 4.4 Given a subset A of a topological space and a point $x \notin \overline{A}$, then $x \notin F$ for some closed set F containing A.

Proof. If $x \notin \overline{A}$, then there is a neighborhood and hence an open set O containing x such that $O \cap A = \emptyset$. Let $F = C(O)$. Then F is closed and $F = C(O) \supset A$. But $x \in O$ and therefore $x \notin F$.

Combining these two lemmas, we obtain,

Theorem 4.5 Given a subset A of a topological space,

$$\overline{A} = \textstyle\bigcap_{\alpha \in I} F_\alpha,$$

where $(F_\alpha)_{\alpha \in I}$ is the family of all closed sets containing A.

Proof. By Lemma 4.3, $\overline{A} \subset \bigcap_{\alpha \in I} F_\alpha$ since $\overline{A} \subset F_\alpha$ for each $\alpha \in I$. By Lemma 4.4, $x \in F_\alpha$ for each $\alpha \in I$ implies that $x \in \overline{A}$, or $\bigcap_{\alpha \in I} F_\alpha \subset \overline{A}$. Thus, $\overline{A} = \bigcap_{\alpha \in I} F_\alpha$.

Frequently, in introducing the concept of closure of a subset, the characterization of closure given by Theorem 4.5 is used as a definition and the statement embodied in our definition, 4.2, is then proved as a theorem. Another possible description of the closure $\overline{A}$ of a subset A, is the characterization of $\overline{A}$ as the smallest closed set containing A. For, on the one hand, $\overline{A}$ is contained in each closed set containing A, whereas on the other hand, $\overline{A}$, being the intersection of closed sets, is itself a closed set.

Theorem 4.5 is the characterization of closure in terms of closed sets. The next theorem characterizes closed sets in terms of closure.

Theorem 4.6 A is closed if and only if $A = \overline{A}$.

Proof. We have just seen that $\overline{A}$ is closed, so if $A = \overline{A}$, then A is closed. Conversely, suppose A is closed. In this event A itself is a closed set containing A, so, therefore, $\overline{A} \subset A$. On the other hand, for an arbitrary subset A, we have $A \subset \overline{A}$, for if $x \in A$, then each neighborhood N of x contains a point of A; namely x itself. Thus, if A is closed, $A = \overline{A}$.

The act of taking the closure of a set associates to each subset A of a topological space a new subset $\overline{A}$. This correspondence or operation on the subsets satisfies the following five properties:

Theorem 4.7 In a topological space $(X, \mathfrak{I})$,

*CL*1. $\overline{\emptyset} = \emptyset$;

*CL*2. $\overline{X} = X$;

*CL*3. For each subset A of X, $A \subset \overline{A}$;

*CL*4. For each pair of subsets A, B of X, $\overline{A \cup B} = \overline{A} \cup \overline{B}$;

*CL*5. For each subset A of X, $\overline{\overline{A}} = \overline{A}$.

Proof. The property *CL*3 has been established during the proof of Theorem 4.6. Note that *CL*2 follows from *CL*3. *CL*1 is true, for given a point $x \in X$ and a neighborhood of N of x, $N \cap \emptyset = \emptyset$; thus there are no points in $\overline{\emptyset}$. To prove *CL*5 we note that $\overline{A}$ is closed, so, applying Theorem 4.6 to $\overline{A}$ we have $\overline{\overline{A}} = \overline{A}$. It remains for us to prove *CL*4. Suppose $x \in \overline{A}$, then each neighborhood N of x contains points of A and hence points of $A \cup B$. Thus $\overline{A} \subset \overline{A \cup B}$. Similarly, $\overline{B} \subset \overline{A \cup B}$, and, consequently, $\overline{A} \cup \overline{B} \subset \overline{A \cup B}$. On the other hand, $A \subset \overline{A}$ and $B \subset \overline{B}$, so $A \cup B \subset \overline{A} \cup \overline{B}$. Thus, $\overline{A} \cup \overline{B}$ is a closed set containing $A \cup B$, whence $\overline{A \cup B} \subset \overline{A} \cup \overline{B}$.

One may use the properties *CL*1–*CL*5 as a set of axioms for what we will call a *closure space* and then prove that there is a "natural" one-one correspondence between the collection of topological spaces and the collection of closure spaces. We shall outline the details of this procedure.

Definition 4.8 Let X be a set. To each subset A of X let there be associated a subset $\overline{A}$ of X, called the closure of A, satisfying the properties *CL*1–*CL*5. The set X together with this set operation of closure is called a *closure space*.

Definition 4.9 In a closure space a subset A is said to be closed if $A = \overline{A}$.

Lemma 4.10 In a closure space, $A \subset B$ implies $\overline{A} \subset \overline{B}$.

Proof. If $A \subset B$, then $A \cup B = B$. Thus

$$\overline{A} \cup \overline{B} = \overline{A \cup B} = \overline{B},$$

whence $\overline{A} \subset \overline{B}$.

Theorem 4.11 In a closure space, the empty set and the whole space are closed. A finite union of closed sets is closed and an arbitrary intersection of closed sets is closed.

Proof. The first three parts follow from *CL*1, *CL*2, and *CL*4, respectively. Finally, if for each $\alpha \in I$, F_α is a closed set so that $\overline{F_\alpha} = F_\alpha$, we have $\bigcap_{\alpha \in I} F_\alpha \subset F_\beta$ for each $\beta \in I$ whence by Lemma 4.10, $\overline{\bigcap_{\alpha \in I} F_\alpha} \subset \overline{F_\beta} = F_\beta$ and therefore $\overline{\bigcap_{\alpha \in I} F_\alpha} \subset \bigcap_{\alpha \in I} F_\alpha$. But by *CL*3, $\bigcap_{\alpha \in I} F_\alpha \subset \overline{\bigcap_{\alpha \in I} F_\alpha}$. Thus, $\bigcap_{\alpha \in I} F_\alpha = \overline{\bigcap_{\alpha \in I} F_\alpha}$ and $\bigcap_{\alpha \in I} F_\alpha$ is closed.

Corollary 4.12 If, in a closure space, a subset is called open provided its complement is closed, the open sets constitute a topology.

Theorem 4.13 In a closure space, for each subset A,

$$\overline{A} = \bigcap_{\alpha \in I} F_\alpha,$$

where $(F_\alpha)_{\alpha \in I}$ is the collection of all closed subsets containing A.

Proof. By *CL*5, $\overline{A}$ is closed and by *CL*3, $\overline{A} \supset A$. Thus $\overline{A} = F_\beta$ for some $\beta \in I$ and $\bigcap_{\alpha \in I} F_\alpha \subset \overline{A}$. But for each $\alpha \in I$, $A \subset F_\alpha$, and by Lemma 4.10, this implies that for each $\alpha \in I$, $\overline{A} \subset \overline{F_\alpha} = F_\alpha$. Thus, $\overline{A} \subset \bigcap_{\alpha \in I} F_\alpha$.

Now, suppose we start with a topological space $(X, \mathfrak{I})$. By Theorem 4.7 this yields a closure space. By Theorem 4.6 the closed subsets of the topological space are precisely the same as the closed subsets of the closure space, hence the same is true of open sets. It follows that the closure space we have constructed gives rise to the original topological space $(X, \mathfrak{I})$. On the other hand, had we started with a closure space and by virtue of Corollary 4.12 defined a top-

ological space, then by comparing Theorems 4.5 and 4.13 we see that the closure operation is the same in both spaces; that is, the topological space gives rise to the original closure space.

In a topological space, we have seen that the closure of a subset A is the smallest closed set containing A. Another significant subset associated with A is the "interior" of A, which, as we shall see, is the largest open set contained in A.

Definition 4.14 Given a subset A of a topological space, a point x is said to be *in the interior of* A if A is a neighborhood of x. The interior of A is denoted by Int (A).

Lemma 4.15 Given a subset A of a topological space and an open set O contained in A,

$$O \subset \text{Int}\,(A).$$

Proof. If $x \in O$, then A is a neighborhood of x, since O is open and $O \subset A$. Thus $x \in \text{Int}\,(A)$ and $O \subset \text{Int}\,(A)$.

Lemma 4.16 Given a subset A of a topological space, if $x \in \text{Int}\,(A)$, then $x \in O$ for some open set $O \subset A$.

Proof. If $x \in \text{Int}\,(A)$, then A is a neighborhood of x, whence A contains an open set O containing x.

In much the same manner in which Lemmas 4.3 and 4.4 combine to yield Theorem 4.5, Lemmas 4.15 and 4.16 combine to yield:

Theorem 4.17 Given a subset A of a topological space,

$$\text{Int}\,(A) = \cup_{\alpha \in I}\, O_\alpha,$$

where $(O_\alpha)_{\alpha \in I}$ is the family of all open sets contained in A.

Thus, Int (A), being the union of open sets, is itself open, and is the largest open set contained in A. Furthermore, if

$(O_\alpha)_{\alpha \in I}$ is the family of open sets contained in a given set A, then $(C(O_\alpha))_{\alpha \in I}$ is the family of closed sets containing $C(A)$. Thus:

Theorem 4.18 $C\,(\text{Int}\,(A)) = \overline{C(A)}$.

Corollary 4.19 $\text{Int}\,(A) = C(\overline{C(A)})$, $C(\overline{A}) = \text{Int}\,(C(A))$.

For a given subset A, the set of points that are arbitrarily close to both A and $C(A)$ is called the "boundary" of A.

Definition 4.20 Given a subset A of a topological space, a point x is said to be *in the boundary of* A if x is in both the closure of A and the closure of the complement of A. The boundary of A is denoted by Bdry (A).

Thus, $\text{Bdry}\,(A) = \overline{A} \cap \overline{C(A)}$. It follows that A and $C(A)$ have the same boundary, for $\text{Bdry}\,C(A) = \overline{C(A)} \cap \overline{C(C(A))} = \overline{C(A)} \cap \overline{A}$. In terms of the definition of the closure of a set, we have the statement that a point x is in the boundary of a set A if and only if each neighborhood N of x contains both points of A and points of the complement of A. Since the boundary of A is the intersection of two closed sets:

Corollary 4.21 For each subset A, **Bdry** (A) is closed.

Theorem 4.22 $\overline{A} = A \cup \text{Bdry}\,(A)$.

Proof. Clearly, $A \subset \overline{A}$ and $\text{Bdry}\,(A) \subset \overline{A}$, so that $\overline{A} \supset \text{Bdry}\,(A) \cup A$. Now, suppose $x \in \overline{A}$. Either $x \in A$ or $x \notin A$. In the event that $x \notin A$, i.e., $x \in C(A)$, $x \in \overline{C(A)}$, whence $x \in \text{Bdry}\,(A) = \overline{A} \cap \overline{C(A)}$. Thus, in either case $x \in A \cup \text{Bdry}\,(A)$ and $\overline{A} \subset A \cup \text{Bdry}\,(A)$.

Theorem 4.23 Let A be a subset of a topological space. A is closed if and only if Bdry $(A) \subset A$. A is open if and only if Bdry $(A) \subset C(A)$.

Proof. If A is closed, Bdry $(A) \subset \overline{A} = A$. Conversely, if Bdry $(A) \subset A$, then $\overline{A} = A \cup$ Bdry $(A) = A$, whence A is closed. Applying what we have just proved to the set $C(A)$, $C(A)$ is closed if and only if Bdry $C(A) \subset C(A)$. But Bdry $C(A) =$ Bdry A. Thus A is open if and only if Bdry $A \subset C(A)$.

Exercises

1. A family $(A_\alpha)_{\alpha \in I}$ of subsets is said to be *mutually disjoint* if for each distinct pair β, γ of indices $A_\beta \cap A_\gamma = \emptyset$. Prove that for each subset A of a topological space $(X, \mathfrak{I})$, the three sets Int (A), Bdry (A), and Int $(C(A))$ are mutually disjoint and that $X =$ Int $(A) \cup$ Bdry $(A) \cup$ Int $(C(A))$.
2. In a metric space (X, d), prove that for each subset A:
   (a) $x \in \overline{A}$ if and only if $d(x, A) = 0$;
   (b) $x \in$ Int (A) if and only if $d(x, C(A)) > 0$;
   (c) $x \in$ Bdry (A) if and only if
   $$d(x, A) = 0 \quad \text{and} \quad d(x, C(A)) = 0.$$
3. In the real line, prove that the boundary of the open interval (a, b) and the boundary of the closed interval $[a, b]$ is $\{a, b\}$.
4. In R^n with the usual topology, let A be the set of points $x = (x_1, x_2, \ldots, x_n)$ such that
   $$x_1^2 + x_2^2 + \ldots + x_n^2 \leqq 1.$$
   Prove that Bdry (A) is the $(n - 1)$-dimensional sphere S^{n-1}, i.e., $x \in$ Bdry (A) if and only if $x_1^2 + x_2^2 + \ldots + x_n^2 = 1$.
5. In R^{n+1} with the usual topology, let A be the set of points $x = (x_1, x_2, \ldots, x_{n+1})$ such that $x_{n+1} = 0$. Prove that Int $(A) = \emptyset$, Bdry $(A) = A$, $\overline{A} = A$.
6. In a topological space, each of the terms *open set, closed set, neighborhood, closure of a set, interior of a set, boundary of a set,* may be characterized by any other one of these terms. Construct a table containing the thirty such possible definitions or theorems in which, for example, the entry in the row labelled interior and the column labelled open set is the characterization of interior in terms of open sets (Theorem 4.17), etc.
7. Let A be a subset of a topological space. Prove that Bdry $(A) = \emptyset$ if and only if A is open and closed.
8. A subset A of a topological space $(X, \mathfrak{I})$ is said to be *dense in* X if $\overline{A} = X$. Prove that if for each open set O we have $A \cap O \neq \emptyset$, then A is dense in X.
9. The "rational density theorem" for the real line states that between any two real numbers there lies a rational number. Use the rational density theorem to prove that the rational numbers are dense in the real line.
10. The "Archimedean principle" for the real line states that if $c, d > 0$ then there is a positive integer N such that $Nc > d$. Prove the Archimedean principle for the real line and use this principle to prove the rational density theorem for the real line.

5 Functions, Continuity, Homeomorphism

Definition 5.1 A *function* f from a topological space $(X, \mathfrak{I})$ to a topological space $(Y, \mathfrak{I}')$ is a function $f: X \to Y$.

If f is a function from a topological space $(X, \mathfrak{I})$ to a topological space $(Y, \mathfrak{I}')$ we shall write $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$. In the event that the topologies on X and Y need not be explicitly mentioned, we may abbreviate this notation by $f: X \to Y$ or simply f.

Definition 5.2 A function $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ is said to be *continuous at a point* $a \in X$ if for each neighborhood N of $f(a)$, $f^{-1}(N)$ is a neighborhood of a. f is said to be *continuous* if f is continuous at each point of X.

Let (X, d) and (Y, d') be metric spaces and let their associated topological spaces be $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$ respectively. Given a function f from the first metric space to the second, we also have a function, which we still denote by f, from the first topological space to the second. Our definition of continuity has been formulated so that for each point $a \in X$, the function $f:(X, d) \to (Y, d')$ is continuous at a if and only if $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ is continuous at a.

Theorem 5.3 A function $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ is continuous if and only if for each open subset O of $Y, f^{-1}(O)$ is an open subset of X.

Proof. First, suppose that f is continuous and that O is an open subset of Y. For each $a \in f^{-1}(O)$, O is a neighborhood of $f(a)$, therefore $f^{-1}(O)$ is a neighborhood of a. Since $f^{-1}(O)$ is a neighborhood of each of its points, $f^{-1}(O)$ is an open subset of X. Conversely, suppose that for each open subset O of $Y, f^{-1}(O)$ is an open subset of X. Let $a \in X$ and a neighborhood N of $f(a)$ be given. N contains an open set O containing $f(a)$, so by our hypothesis, $f^{-1}(N)$ contains the open set $f^{-1}(O)$ containing a. Thus, $f^{-1}(N)$ is a neighborhood of a and f is continuous at a. Since a was arbitrary, f is continuous.

It is important to remember that Theorem 5.3 says that a function f is continuous if and only if the *inverse* image of each open set is open. This characterization of continuity should not be confused with another property that a function may or may not possess, the property that the image of each open set is an open set (such functions are called *open mappings.*) There are many situations in which a function $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ has the property that for each open subset A of X, the set $f(A)$ is an open subset of Y, and yet f is *not* continuous. For example, let Y be a set containing two distinct elements a and b and let each subset of Y be an open set. Define $f:R \to Y$ by $f(x) = a$ for $x \geqq 0$ and $f(x) = b$ for $x < 0$. Then for each open subset O of R, $f(O)$ is an open subset of Y, whereas f is not continuous at $0 \in R$, for $\{a\}$ is a neighborhood of $f(0)$ whereas $f^{-1}(\{a\})$ is not a neighborhood of 0.

Theorem 5.4 A function $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ is continuous if and only if for each closed subset F of $Y, f^{-1}(F)$ is a closed subset of X.

Proof. F is closed if and only if $C(F)$ is open. But $f^{-1}(C(F)) = C(f^{-1}(F))$. Therefore f is continuous implies that $f^{-1}(C(F))$ is open whenever F is closed, or that $f^{-1}(F)$ is closed. Conversely, if for each closed subset F of $Y, f^{-1}(F)$ is a closed subset of X, then, given an open subset O of Y, $f^{-1}(C(O)) = C(f^{-1}(O))$ is closed, hence $f^{-1}(O)$ is open and f is continuous.

Theorem 5.5 $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ is continuous if and only if for each subset A of X,

$$f(\overline{A}) \subset \overline{f(A)}.$$

Proof. First, suppose that f is continuous. For each subset A of X, $A \subset f^{-1}(f(A))$. Now $f(A) \subset \overline{f(A)}$, and therefore $A \subset f^{-1}(f(A)) \subset f^{-1}(\overline{f(A)})$. But $\overline{f(A)}$ is closed and since f is continuous, $f^{-1}(\overline{f(A)})$ is closed. Thus $f^{-1}(\overline{f(A)})$ is a closed set containing A, whence

$$\overline{A} \subset f^{-1}(\overline{f(A)}),$$

from which it follows that $f(\overline{A}) \subset f(f^{-1}(\overline{f(A)})) = \overline{f(A)}$. Conversely, suppose that for each subset A of X, $f(\overline{A}) \subset \overline{f(A)}$. Let F be a closed subset of Y. To show that f is continuous it suffices to show that $f^{-1}(F)$ is closed; that is, $f^{-1}(F) = \overline{f^{-1}(F)}$. By our hypothesis $f(\overline{f^{-1}(F)}) \subset \overline{f(f^{-1}(F))} = \overline{A} = F$. Thus (applying f^{-1} to this last relation) $\overline{f^{-1}(F)} = f^{-1}(f(\overline{f^{-1}(F)})) \subset f^{-1}(F)$. On the other hand, $f^{-1}(F) \subset \overline{f^{-1}(F)}$, so that $f^{-1}(F) = \overline{f^{-1}(F)}$ and f is continuous.

Theorem 5.6 Let $f:(X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ be continuous at a point $a \in X$ and let $g:(Y, \mathfrak{I}') \to (Z, \mathfrak{I}'')$ be continuous at $f(a)$. Then the composite function $gf:(X, \mathfrak{I}) \to (Z, \mathfrak{I}'')$ is continuous at a.

Proof. Let N be a neighborhood of $(gf)(a) = g(f(a))$. Then $(gf)^{-1}(N) = f^{-1}(g^{-1}(N))$. But $g^{-1}(N)$ is a neighborhood of $f(a)$, since g is continuous at $f(a)$, and therefore $f^{-1}(g^{-1}(N))$ is a neighborhood of a, since f is continuous at a.

The equivalence relation that is appropriate to topological spaces is called *homeomorphism*.

Definition 5.7 Topological spaces $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$ are called *homeomorphic* if there exist inverse functions $f: X \to Y$ and $g: Y \to X$ such that f and g are continuous. In this event the functions f and g are said to be *homeomorphisms* and we say that f and g define a homeomorphism between $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$.

The following easily verified corollary to this definition indicates that homeomorphism is the translation from metric spaces to topological spaces of the concept of topological equivalence.

Corollary 5.8 Let (X, d) and (Y, d') be metric spaces. Let $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$ be the topological spaces associated with (X, d) and (Y, d') respectively. Then the metric spaces (X, d) and (Y, d') are topologically equivalent if and only if the topological spaces $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$ are homeomorphic.

Theorem 5.9 A necessary and sufficient condition that two topological spaces $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$ be homeomorphic is that there exist a function $f: X \to Y$ such that:

1. f is one-one;
2. f is onto;
3. A subset O of X is open if and only if $f(O)$ is open.

Proof. Suppose that $(X, \mathfrak{I})$ and $(Y, \mathfrak{I}')$ are homeomorphic. Let the homeomorphism be defined by inverse functions $f: X \to Y$ and $g: Y \to X$. f is invertible and consequently one-one and onto. Furthermore, given an open set O in X, the set $f(O) = g^{-1}(O)$ is open in Y, since g is continuous. On the other hand, if O' is an open subset of Y, then $O = f^{-1}(O')$ is open in X and $f(O) = O'$.

Now, suppose that a function $f: X \to Y$ with the prescribed properties exists. Then f is invertible and we define $g: Y \to X$ by $g(b) = a$ if $f(a) = b$, so that f and g are inverse functions. If O is an open subset of X, then $f(O) = g^{-1}(O)$ is open in Y, so that g is continuous.

Also, if O' is an open subset of Y, then $f^{-1}(O') = O$ is an open subset of X and f is continuous.

Exercises

1. Let X be an arbitrary set. Prove that the identity mapping
$$i: (X, 2^X) \to (X, \{\emptyset, X\})$$
is continuous. Prove that if X has more than one distinct element, then the identity mapping
$$i: (X, \{\emptyset, X\}) \to (X, 2^X)$$
is not continuous.

2. Let X be an arbitrary set. Prove that the identity mapping
$$i: (X, \mathfrak{I}) \to (X, \mathfrak{I}')$$
is continuous if and only if $\mathfrak{I}' \subset \mathfrak{I}$.

3. Prove that a function $f: (X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ is a homeomorphism if and only if
   (i) f is one-one;
   (ii) f is onto;
   (iii) For each point $x \in X$ and each subset N of X, N is a neighborhood of x if and only if $f(N)$ is a neighborhood of $f(x)$.

4. Let $f: (X, \mathfrak{I}) \to (Y, \mathfrak{I}')$ be a homeomorphism. Let a third topological space $(Z, \mathfrak{I}'')$ and a function $h: (Y, \mathfrak{I}') \to (Z, \mathfrak{I}'')$ be given. Prove that h is continuous if and only if hf is continuous. Let another function $k: (Z, \mathfrak{I}'') \to (X, \mathfrak{I})$ be given. Prove that k is continuous if and only if fk is continuous.

5. Prove that the function $f: R \to R$ defined by $f(x) = \sin x$ is continuous. $\left[\textit{Hint: } |\sin a - \sin x| = 2\left|\sin \frac{a-x}{2} \cos \frac{a+x}{2}\right| \text{ and } \left|\sin \frac{a-x}{2}\right| \leq \left|\frac{a-x}{2}\right|.\right]$ Find an open interval (a, b) such that $f((a, b))$ is *not* an open interval.

6 Subspaces

Definition 6.1 Let $(X, \mathfrak{J})$ and $(Y, \mathfrak{J}')$ be topological spaces. The topological space Y is called a *subspace* of the topological space X if $Y \subset X$ and if each open subset O' of Y is of the form

$$O' = O \cap Y$$

for some open subset O of X.

In the event that Y is a subspace of X, we may say that each open subset O' of Y is the restriction to Y of an open subset O of X. A subset O' that is open in Y is often called *relatively open in* Y or simply *relatively open*. A subset O of X that is open in X and is contained in Y is necessarily relatively open in Y, but the relatively open subsets of Y are in general not open in X.

We shall now prove that there are as many subspaces of a topological space X as there are subsets Y of X.

Proposition 6.2 Let $(X, \mathfrak{J})$ be a topological space and let Y be a subset of X. Define the collection $\mathfrak{J}'$ of subsets of Y as the collection of subsets O' of Y of the form

$$O' = O \cap Y,$$

where $O \in \mathfrak{J}$. Then $(Y, \mathfrak{J}')$ is a topological space and therefore a subspace of $(X, \mathfrak{J})$.

Proof. We must prove that $\mathfrak{J}'$ is a topology. $\emptyset = \emptyset \cap Y$ and $Y = X \cap Y$ are in $\mathfrak{J}$. Suppose that $O_1', O_2', \ldots, O_n' \in \mathfrak{J}'$, so that for $i = 1, 2, \ldots, n$,

$$O_i' = O_i \cap Y$$

for some $O_i \in \mathfrak{J}$. Then

$$O_1' \cap O_2' \cap \cdots \cap O_n' = (O_1 \cap O_2 \cap \cdots \cap O_n) \cap Y$$

is in $\mathfrak{J}'$, since $O_1 \cap O_2 \cap \cdots \cap O_n$ is open in X. Finally, suppose that for each $\alpha \in I$, $O_\alpha' \in \mathfrak{J}'$. Thus, for each $\alpha \in I$,

$$O_\alpha' = O_\alpha \cap Y$$

for some $O_\alpha \in \mathfrak{J}$. But

$$\cup_{\alpha\in I} O_\alpha' = \cup_{\alpha\in I} (O_\alpha \cap Y) = (\cup_{\alpha\in I} O_\alpha) \cap Y$$

is in $\mathfrak{J}'$, since $\cup_{\alpha\in I} O_\alpha$ is open in X.

Given a subset Y of a topological space $(X, \mathfrak{J})$, the topology $\mathfrak{J}'$ of Y described in the above proposition is said to be *induced* by the topology $\mathfrak{J}$ on X and is called the *relative topology* on Y. The neighborhoods in this relative topology on Y are called *neighborhoods in* Y or *relative neighborhoods*. The following result states that the neighborhoods in Y are the restrictions of the neighborhoods in X.

Theorem 6.3 Let Y be a subspace of a topological space X and let $a \in Y$. Then a subset N' of Y is a relative neighborhood of a if and only if

$$N' = N \cap Y,$$

where N is a neighborhood of a in X.

Proof. First suppose N' is a relative neighborhood of a. Then N' contains a relatively open set O', which contains a. Let $O' = O \cap Y$, where O is an open subset of X. Then $N = N' \cup O$ is a neighborhood of a in X and $N \cap Y = (N' \cup O) \cap Y = N' \cup (O \cap Y) = N'$. Conversely, suppose $N' = N \cap Y$, where N is a neighborhood of a in X. Then N contains an open set O containing a and hence N' contains the relatively open set $O' = O \cap Y$ containing a. Thus N' is a relative neighborhood of a.

Example 1 The closed interval $[a, b]$ of the real line with induced topology is a subspace of the real line. A relative neighborhood of the point a is any subset N of $[a, b]$ that contains a half-open interval $[a, c)$, where $a < c$ and $[a, c)$ is the set of all real numbers x such that $a \leqq x < c$. Similarly, a relative neighborhood of the point b is any subset M of $[a, b]$ that contains a half-open interval $(c, b]$, where $c < b$ and $(c, b]$ is the set of all real numbers x such that $c < x \leqq b$. If d is such that $a < d < b$, then a relative neighborhood of d is any subset U of $[a, b]$ that is a neighborhood of d in the real line R.

Example 2 Let A be the subset of R^{n+1} consisting of all points $x = (x_1, x_2, \ldots, x_{n+1})$ such that $x_{n+1} = 0$. Let R^{n+1} have the usual topology and let A have the induced topology so that A is a subspace of R^{n+1}. The topological space A is homeomorphic to R^n. To prove this fact we shall use the result that the relationship of subspace is "preserved" in passing from metric spaces to topological spaces.

Lemma 6.4 Let (X, d) be a metric space and let (Y, d') be a subspace of (X, d). If $(X, \mathfrak{I})$ is the topological space associated with (X, d) and $(Y, \mathfrak{I}')$ is the topological space associated with (Y, d'), then $(Y, \mathfrak{I}')$ is a subspace of $(X, \mathfrak{I})$.

Proof. By Theorem 9.5, Chapter II, a subset O' of the metric space (Y, d') is open if and only if there is an open subset O of the metric space (X, d) such that $O' = O \cap Y$. Thus, the topology $\mathfrak{I}'$ on Y consists of the restriction to Y of the open sets $O \in \mathfrak{I}$, and $(Y, \mathfrak{I}')$ is a subspace of $(X, \mathfrak{I})$.

Returning to our example, we define $f: R^n \rightarrow A$ by setting $f(x_1, x_2, \ldots, x_n) = (x_1, x_2, \ldots, x_n, 0)$. It is easily verified that f is one-one, onto, and that the inverse of f is the function $g: A \rightarrow R^n$ defined by $g(x_1, x_2, \ldots, x_n, 0) = (x_1, x_2, \ldots, x_n)$. If we first consider f and g as functions defined on the metric spaces (R^n, d) and (A, d'), where (A, d') is a subspace of (R^{n+1}, d), then clearly f and g are continuous. It follows that f and g are continuous functions defined on the topological spaces R^n and A, where A is considered as a subspace of R^{n+1}, and that they therefore define a homeomorphism.

Given a subspace $(Y, \mathfrak{I}')$ of a topological space $(X, \mathfrak{I})$, the closed subsets of the topological space $(Y, \mathfrak{I}')$ are called *relatively closed in* Y or simply *relatively closed*. Again, the relatively closed subsets are the restriction to Y of the closed subsets of X.

Theorem 6.5 Let $(Y, \mathfrak{I}')$ be a subspace of the topological space $(X, \mathfrak{I})$. A subset F' of Y is relatively closed in Y if and only if

$$F' = F \cap Y,$$

for some closed subset F of X.

Proof. First, suppose F' is relatively closed. Then $C_Y(F')$ is relatively open. Thus, $C_Y(F') = O \cap Y$, where O is open in X. But then $F' = C_Y(O \cap Y) = C_Y(O) = C_X(O) \cap Y$, where $C_X(O)$ is a closed subset of X. Conversely, suppose $F' = F \cap Y$, where F is a closed subset of X. Then, $C_Y(F') = C_X(F) \cap Y$; hence $C_Y(F')$ is relatively open in Y and therefore F' is relatively closed.

Example 3 Let $a < b < c < d$. Let $Y = [a, b] \cup (c, d)$ be considered as a subspace of the real line. Then the subset $[a, b]$ of Y is both relatively open and relatively closed. To prove this fact we note that $[a, b] = [a, b] \cap Y$ so that $[a, b]$ is relatively closed, whereas for $0 < \varepsilon < c - b$, $[a, b] = (a - \varepsilon, b + \varepsilon) \cap Y$ so that $[a, b]$ is relatively open. Since (c, d) is the complement in Y of a relatively open and relatively closed subset of Y, (c, d) is also relatively open and relatively closed in Y.

Theorem 6.6 Let the topological space Y be a subspace of the topological space X. Then the inclusion mapping $i: Y \rightarrow X$ is continuous.

Proof. For each subset A of X, $i^{-1}(A) = A \cap Y$. Thus, if O is an open subset of X, $i^{-1}(O) = O \cap Y$ is a relatively open subset of Y.

Exercises

1. If Y is a subspace of X and Z is a subspace of Y, then Z is a subspace of X.
2. Let O be an open subset of a topological space X. Prove that a subset A of O is relatively open if and only if it is an open subset of X.
3. Let F be a closed subset of a topological space X. Prove that a subset A of F is relatively closed if and only if it is a closed subset of X.
4. Prove that a subspace of a Hausdorff space is a Hausdorff space.
5. Prove that a subspace of a metrizable space is a metrizable space.

6. Prove that an open interval (a, b) considered as a subspace of the real line is homeomorphic to the real line.

7. Let Y be a subspace of X and let A be a subset of Y. Denote by $\text{Int}_X(A)$ the interior of A in the topological space X and by $\text{Int}_Y(A)$ the interior of A in the topological space Y. Prove that $\text{Int}_X(A) \subset \text{Int}_Y(A)$. Illustrate by an example the fact that in general $\text{Int}_X(A) \neq \text{Int}_Y(A)$.

8. Let Y be a subspace of X and let A be a subset of Y. Denote by $\bar{A}^X$ the closure of A in the topological space X and by $\bar{A}^Y$ the closure of A in the topological space Y. Prove that $\bar{A}^Y \subset \bar{A}^X$. Show that in general $\bar{A}^Y \neq \bar{A}^X$.

7 Products

Throughout this section let $(X_1, \mathfrak{I}_1), (X_2, \mathfrak{I}_2), \ldots, (X_n, \mathfrak{I}_n)$ be topological spaces and let $X = \prod_{i=1}^{n} X_i$. We wish to define a topology on X that may be regarded as the product of the topologies on the factors of X. Again our guide is the corresponding situation in metric spaces. If these topological spaces were metrizable, then there is a standard procedure for converting the product of the corresponding metric spaces into a metric space. In this resulting metric space, the open subsets of X are the unions of sets of the form $O_1 \times O_2 \times \ldots \times O_n$, where each O_i is an open subset of X_i. If, in general, we use these subsets of the product set X to construct a topology on X, we must verify that we do, in fact, have a topology.

Lemma 7.1 Let $\mathfrak{I}$ be the collection of subsets of $X = \prod_{i=1}^{n} X_i$, which are unions of sets of the form

$$O_1 \times O_2 \times \ldots \times O_n,$$

where each O_i is an open subset of X_i. Then $\mathfrak{I}$ is a topology.

Proof. Clearly $\emptyset$ and X are in $\mathfrak{I}$. Suppose O and O' are in $\mathfrak{I}$. Then

$$O = \cup_{\alpha \in I} O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n},$$

$$O' = \cup_{\beta \in J} O'_{\beta,1} \times O'_{\beta,2} \times \ldots \times O'_{\beta,n},$$

where for each $\alpha \in I$, $O_{\alpha,i}$ is an open subset of X_i and for each $\beta \in J$, $O'_{\beta,i}$ is an open subset of X_i. For each $(\alpha, \beta) \in I \times J$ and each $i = 1, 2, \ldots, n$, set

$$O_{(\alpha,\beta),i} = O_{\alpha,i} \cap O'_{\beta,i}.$$

Thus, $O_{(\alpha,\beta),i}$ is an open subset of X_i and

$$O \cap O' = \cup_{(\alpha,\beta) \in I \times J} O_{(\alpha,\beta),1} \times O_{(\alpha,\beta),2} \times \ldots \times O_{(\alpha,\beta),n}$$

is also in $\mathfrak{I}$. Finally, a union of sets, each of which is a union of sets of this specified form is again a union of sets of this specified form, and consequently $\mathfrak{I}$ is a topology.

Definition 7.2 The topological space $(X, \mathfrak{I})$, where $\mathfrak{I}$ is the collection of subsets of X that are unions of sets of the form

$$O_1 \times O_2 \times \ldots \times O_n,$$

each O_i an open subset of X_i, is called the *product* of the topological spaces $(X_i, \mathfrak{I}_i)$, $i = 1, 2, \ldots, n$.

In the future we shall often denote a topological space $(X, \mathfrak{I})$ simply by X. Thus, if we say, let $X_1, X_2, \ldots, X_n$ be topological spaces and $X = \prod_{i=1}^{n} X_i$, we shall mean that X is to be considered as the product of the topological spaces.

As we did with metric spaces, we have used the sets of the form $O_1 \times O_2 \times \ldots \times O_n$, O_i open in X_i, as a "basis" for the open sets of X.

Definition 7.3 Let X be a topological space and $(O_\alpha)_{\alpha \in I}$ a collection of open sets in X. $(O_\alpha)_{\alpha \in I}$ is called a *basis for the open sets of X* if each open set is a union of members of $(O_\alpha)_{\alpha \in I}$.

The next proposition characterizes the neighborhoods in the product space.

Proposition 7.4 In a topological space $X = \prod_{i=1}^{n} X_i$, a subset N is a neighborhood of a point $a = (a_1, a_2, \ldots, a_n) \in N$ if and only if N contains a subset of the form $N_1 \times N_2 \times \ldots \times N_n$, where each N_i is a neighborhood of a_i.

Proof. First suppose that $N_1 \times N_2 \times \ldots \times N_n \subset N$ where each N_i is a neighborhood of a_i. By the definition of neighborhood in a topological space, each N_i contains an open set O_i containing a_i, hence N contains the open set $O_1 \times O_2 \times \ldots \times O_n$ containing a, and therefore N is a neighborhood of a. Conversely, suppose N is a neighborhood of a. Then N contains an open set O containing a. Since O is an open subset of the product space $X = \prod_{i=1}^{n} X_i$, we may write $O = \bigcup_{\alpha \in I} O_{\alpha,1} \times O_{\alpha,2} \times \ldots \times O_{\alpha,n}$, where for each i and each $\alpha \in I$, $O_{\alpha,i}$ is an open subset of X_i. Since $a \in O$, $a \in O_{\beta,1} \times O_{\beta,2} \times \ldots \times O_{\beta,n}$, for some $\beta \in I$, hence $a_i \in O_{\beta,i}$ for $i = 1, 2, \ldots, n$. But $O_{\beta,i}$ is open. Thus, if we set $N_i = O_{\beta,i}$, $i = 1, 2, \ldots, n$, N_i is a neighborhood of a_i and $N_1 \times N_2 \times \ldots \times N_n \subset O \subset N$.

Definition 7.5 Let X be a topological space and $a \in X$. A collection $\mathfrak{N}$ of neighborhoods of a is called a *basis for the neighborhoods at a* if each neighborhood N of a contains a member of $\mathfrak{N}$.

Thus, if $a = (a_1, a_2, \ldots, a_n) \in X = \prod_{i=1}^{n} X_i$, a basis for the neighborhoods at a is the collection consisting of all subsets of the form $N_1 \times N_2 \times \ldots \times N_n$, where each N_i is a neighborhood of a_i.

Exercises

1. Prove that a subset F of $X = \prod_{i=1}^{n} X_i$ is closed if and only if F is an intersection of sets, each of which is a finite union of sets of the form $F_1 \times F_2 \times \ldots \times F_n$, where each F_i is a closed subset of X_i.

2. Let $X = \prod_{i=1}^{n} X_i$ be a product of topological spaces and let $a = (a_1, a_2, \ldots, a_n) \in X$. For each $i = 1, 2, \ldots, n$, let $\mathfrak{N}_i$ be a basis for the neighborhoods at a_i. Prove that the collection $\mathfrak{N}$ of subsets of X of the form $N_1 \times N_2 \times \ldots \times N_n$, where each $N_i \in \mathfrak{N}_i$, is a basis for the neighborhoods at a.

3. Define the i^{th} *projection* $p_i: X \to X_i$ for $i = 1, 2, \ldots, n$ by $p_i(x_1, x_2, \ldots, x_n) = x_i$. Prove that each projection is a continuous function. Let Y be a set. If Y is a topological space, prove that a function $f: Y \to X$ is continuous if and only if the n functions $p_i f$ are continuous.

4. Prove that the function $f: R \to R^2$ defined by

$$f(x) = (\cos 2\pi x, \sin 2\pi x)$$

is continuous. What is $f^{-1}(\{(1, 0)\})$?

Connectedness

IV

1 Introduction

A subspace of a topological space is "connected" if it is all "of one piece"; that is, if it is impossible to decompose the subspace into two disjoint non-empty open sets. The non-empty connected subsets of the real line are single points and intervals. The continuous image of a connected set is necessarily a connected set. A consequence of these two facts is the intermediate value theorem; that is, a continuous function $f:[a, b] \to R$ must assume all values between $f(a)$ and $f(b)$. A second type of connectedness is called "arcwise connectedness," by which it is meant that each pair of points may be "connected" by a "path" or "arc." Arcwise connectedness is a stronger condition than connectedness, since each arcwise connected topological space is connected, whereas the converse is false. A third type of connectedness that we shall consider

is "simple connectedness." A topological space is simply connected if there are no holes in it to prevent the continuous shrinking of each closed arc to a point.

2 Connectedness

Definition 2.1 A topological space X is said to be *connected* if the only two subsets of X that are simultaneously open and closed are X itself and the empty set $\emptyset$.

Thus, a topological space X is not connected if and only if there are two non-empty open subsets P and Q whose union is X and whose intersection is empty, for in this event P is the complement of Q and therefore both open and closed, whereas P is neither X nor $\emptyset$. Similarly, a topological space X is not connected if and only if there are two non-empty closed subsets F and G whose union is X and whose intersection is empty.

Every subset A of a topological space X is itself a topological space in the relative topology. We say that the subset A is connected if the topological space A with the relative topology is connected, or what amounts to the same thing,

Definition 2.2 A subset A of a topological space X is said to be *connected* if the only two subsets of A that are simultaneously relatively open and relatively closed in A are A and $\emptyset$.

Thus, the statement, A is connected, has the same meaning whether the reference is to A as a topological space or as a subspace of some larger topological space.

We shall shortly see that intervals such as $[a, b]$ and (a, b) are connected subsets of the real line R. As an example of a subset of the real line that is not connected, let

$A = [0, 1] \cup (2, 3)$. $[0, 1]$ is a relatively closed subset of A since $[0, 1]$ is closed in R. At the same time $[0, 1]$ is a relatively open subset of A, since $[0, 1] = (-\frac{1}{2}, \frac{3}{2}) \cap A$. Finally, $[0, 1] \neq \emptyset$ and $[0, 1] \neq A$, hence A is not connected. By the same token, the "open interval" $(2, 3)$ is also both relatively open and relatively closed in A.

It will be useful to have the following formulation of connectedness, or more precisely, non-connectedness.

Lemma 2.3 Let A be a subspace of a topological space X. Then A is not connected if and only if there exist two open subsets P and Q of X such that,

$$A \subset P \cup Q,$$
$$P \cap Q \subset C(A),$$

and $P \cap A \neq \emptyset$, $Q \cap A \neq \emptyset$.

Proof. First, suppose that A is not connected. Then there is a subset P' of A that is different from $\emptyset$ and from A and is both relatively open and relatively closed. This implies that the complement of P' in A, $C_A(P')$, is also different from $\emptyset$ and from A and relatively open. Thus $P' = P \cap A$ and $C_A(P') = Q \cap A$, where P and Q are open subsets of X. We therefore have that $A = P' \cup C_A(P') \subset P \cup Q$, for $P' \subset P$ and $C_A(P') \subset Q$, and also $P \cap Q \cap A = (P \cap A) \cap (Q \cap A) = P' \cap C_A(P') = \emptyset$ so that $P \cap Q \subset C(A)$. Finally, $P' = P \cap A$ and $C_A(P') = Q \cap A$ are non-empty.

Conversely, given open sets P and Q satisfying the stated conditions, set $P' = P \cap A$ and $Q' = Q \cap A$. Then $A = A \cap (P \cup Q) = (A \cap P) \cup (A \cap Q) = P' \cup Q'$ and $P' \cap Q' = (A \cap P) \cap (A \cap Q) = \emptyset$. Thus $P' = C_A(Q')$, and P' is both relatively open and relatively closed in A. Since $P' \neq \emptyset$ and $P' \neq A$ (for Q' is non-empty), A is not connected.

A corresponding result also holds, using closed sets.

Lemma 2.4 Let A be a subspace of a topological space X. Then A is not connected if and only if there exist two closed subsets F and G of X such that

$$A \subset F \cup G,$$
$$F \cap G \subset C(A),$$

and $F \cap A \neq \emptyset$, $G \cap A \neq \emptyset$.

The next theorem asserts that connectedness is preserved under continuous mappings.

Theorem 2.5 Let X and Y be topological spaces and let $f: X \to Y$ be continuous. If A is a connected subset of X, then $f(A)$ is a connected subset of Y.

Proof. Suppose $f(A)$ is not connected. Then there are open subsets P' and Q' of Y such that $f(A) \subset P' \cup Q'$, $P' \cap Q' \subset C(f(A))$, and $P' \cap f(A) \neq \emptyset$, $Q' \cap f(A) \neq \emptyset$. Since f is continuous, $P = f^{-1}(P')$ and $Q = f^{-1}(Q')$ are open subsets of X. But $A \subset f^{-1}(f(A)) \subset f^{-1}(P' \cup Q') = P \cup Q$. Also $P \cap Q = f^{-1}(P' \cap Q') \subset f^{-1}(Cf(A)) = C(f^{-1}(f(A))) \subset C(A)$. Finally, $P \cap A \neq \emptyset$, $Q \cap A \neq \emptyset$. Thus, A is not connected. It follows that if A is connected then $f(A)$ must also be connected.

In particular, if X is connected and $f: X \to Y$ is continuous, then $f(X)$ is connected. If, under these circumstances, we wish to assert that Y is connected, we must require that $f(X) = Y$; that is, that f is onto.

Corollary 2.6 Let X and Y be topological spaces, let $f: X \to Y$ be a continuous mapping of X onto Y, and let X be connected; then Y is connected.

Corollary 2.7 Let X and Y be homeomorphic topological spaces, then X is connected if and only if Y is connected.

A property of a topological space is said to be a *topological property* if each topological space homeomorphic to the given space must also possess this property. Thus, Corollary 2.7 states that connectedness is a topological property. Of course,

any property that is preserved under continuous mappings must necessarily be a topological property.

Lemma 2.8 supplies an interesting characterization of connectedness, which will facilitate our proving that the product of two connected spaces is itself connected.

Lemma 2.8 Let $Y = \{0, 1\}$. A topological space X is connected if and only if the only continuous mappings $f: X \to Y$ are the constant mappings.

Proof. Let $f: X \to Y$ be a continuous non-constant mapping. Then $P = f^{-1}(\{0\})$ and $Q = f^{-1}(\{1\})$ are both non-empty. Thus, $P \neq \emptyset$ and $P \neq X$. $\{0\}$ and $\{1\}$ are open subsets of Y and f is continuous, therefore P and Q are open subsets of X. But $P = C(Q)$, so P is both open and closed and consequently X is not connected. Thus, if X is connected, the only continuous mappings $f: X \to Y$ are constant mappings.

Conversely, suppose X is not connected. Then there are non-empty open subsets P, Q of X such that $P \cap Q = \emptyset$ and $P \cup Q = X$. Define a mapping $f: X \to Y$ as follows. If $x \in P$, set $f(x) = 0$. If $x \in Q$, set $f(x) = 1$. f is continuous, for there are four open subsets, $\emptyset$, $\{0\}$, $\{1\}$, and Y of Y whereas $f^{-1}(\emptyset) = \emptyset$, $f^{-1}(\{0\}) = P$, $f^{-1}(\{1\}) = Q$, and $f^{-1}(Y) = X$, so that the inverse image of an open set is open.

Clearly, the role of the space $Y = \{0, 1\}$ in the above result could be played by any other topological space Z consisting of two points in which all subsets are open.

Theorem 2.9 Let X and Y be connected topological spaces. Then $X \times Y$ is connected.

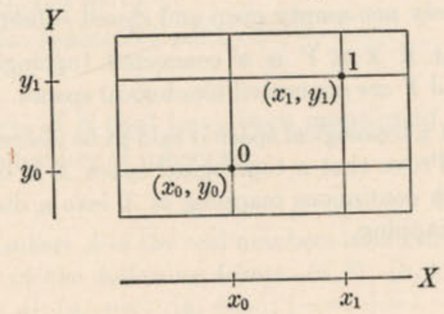
Proof. We shall show that the only continuous mappings $f: X \times Y \to \{0, 1\}$ are constant mappings. Suppose, on the contrary, that there is a continuous mapping $f: X \times Y \to \{0, 1\}$ that is not constant. Then there are points (x_0, y_0), $(x_1, y_1) \in X \times Y$ such that $f(x_0, y_0) = 0$, $f(x_1, y_1) = 1$. If we picture $f(x, y)$ as a number attached to the point (x, y), then we have the situation depicted in

Figure 12. Suppose $f(x_1, y_0) = 0$. We then define an "imbedding" $i_{x_1}: Y \to X \times Y$ by $i_{x_1}(y) = (x_1, y)$. i_{x_1} is continuous, hence the composite mapping

$$fi_{x_1}: Y \to \{0, 1\}$$

is continuous (fi_{x_1} may be thought of as essentially f restricted to the points of the form (x_1, y); that is, the points lying above x_1 in Figure 12.) But $(fi_{x_1})(y_0) = f(x_1, y_0) = 0$ and $(fi_{x_1})(y_1) = f(x_1, y_1) = 1$.

Figure 12



Thus, in this case, there is a non-constant mapping of Y into $\{0, 1\}$, contradicting the connectedness of Y. Similarly, if $f(x_1, y_0) = 1$, we define an imbedding $i_{y_0}: X \to X \times Y$ by setting $i_{y_0}(x) = (x, y_0)$ and show that we obtain a non-constant mapping $fi_{y_0}: X \to \{0, 1\}$, contradicting the connectedness of X. It follows that there are no non-constant mappings of $X \times Y$ into $\{0, 1\}$ and that therefore $X \times Y$ is connected.

Corollary 2.10 If $X_1, X_2, \ldots, X_n$ are connected topological spaces, then $\prod_{i=1}^{n} X_i$ is a connected topological space.

Exercises

1. On the real line, prove that the set of non-zero numbers is not a connected set.

2. On the real line, prove that the set of rational numbers is not a connected set.
3. Let A and B be subsets of a topological space X. If A is connected, B is open and closed, and $A \cap B \neq \emptyset$, prove that $A \subset B$. [*Hint:* Assume $A \not\subset B$ and use the sets $P = A \cap B$ and $Q = A \cap C(B)$ to prove that A is not connected.]
4. Let A and B be connected subsets of a topological space X. If $A \cap B \neq \emptyset$, prove that $A \cup B$ is connected. [*Hint:* in the topological space $A \cup B$, show by using the result of Problem 3 that the only non-empty open and closed subset is $A \cup B$.]
5. Prove that if $X \times Y$ is a connected topological space, then both X and Y are connected topological spaces.
6. Recall that a topological space is said to be discrete if each subset is open. Prove that a topological space X is connected if and only if each continuous mapping of X into a discrete space is a constant mapping.

3 Connectedness on the Real Line

In this section we shall define the term "interval" and prove that a non-empty subset of the real line is connected if and only if it is either a single point or an interval.

Definition 3.1 A subset A of the real line is called an *interval* if A contains at least two distinct points, and if given points $a, b \in A$ with $a < b$, then for each point x such that $a < x < b$, it follows that $x \in A$.

Thus, an interval contains all points between any two of its points. It is a simple matter to verify that a closed interval $[a, b]$ or an open interval (a, b) is an interval in the sense of Definition 3.1. Other subsets of the real line that are intervals are defined in Definition 3.2.

Definition 3.2 Let a be a real number. The subset of R consisting of all real numbers x such that $a < x$ is denoted by $(a, +\infty)$. The subset of R consisting of all real numbers x such that $a \leqq x$ is denoted by $[a, +\infty)$. The subset of R consisting of all real numbers x such that $x < a$ is denoted by $(-\infty, a)$. The subset of R consisting of all real numbers x such that $x \leqq a$ is denoted by $(-\infty, a]$.

Let b be a second real number with $a < b$. The subset of R consisting of all real numbers x such that $a < x \leqq b$ is denoted by $(a, b]$. The subset of R consisting of all real numbers x such that $a \leqq x < b$ is denoted by $[a, b)$.

We shall also denote R itself by $(-\infty, +\infty)$.

The subsets of R that have been mentioned in this section exhaust the collection of intervals.

Theorem 3.3 A subset A of the real numbers is an interval if and only if it is of one of the following forms: (a, b); $[a, b)$; $(a, b]$; $[a, b]$; $(-\infty, a)$; $(-\infty, a]$; $(a, +\infty)$; $[a, +\infty)$; $(-\infty, +\infty)$.

Proof. We leave it to the reader to verify that each of these nine types of sets is an interval and shall prove the "only if" part of the theorem. Suppose A is an interval. We first note that if a point $x \notin A$, then either x is a lower bound of A or an upper bound of A, for otherwise there would be points $a, b \in A$ with $a < x < b$ and we would obtain the contradiction $x \in A$. We shall, consequently, distinguish four cases.

Case 1. A has neither an upper bound nor a lower bound. In this case $C(A)$ must be empty so that $A = (-\infty, +\infty)$.

Case 2. A has an upper bound but no lower bound. Since an interval is non-empty, A has a least upper bound a. We claim that if $x < a$, then $x \in A$. For, suppose $x < a$, then there is a point $a' \in A$ with $x < a' \leqq a$ (for otherwise a would not be a least upper bound.) Since x cannot be a lower bound of A there is a point $b \in A$ with $b < x$. But $b < x < a'$ and $a', b \in A$ imply that $x \in A$. We have thus shown that $(-\infty, a) \subset A$. On the other hand, for $x > a$, $x \notin A$. It follows that A is either of the form $(-\infty, a]$ or $(-\infty, a)$, depending on whether $a \in A$ or $a \notin A$.

Case 3. A has a lower bound but no upper bound. By reasoning similar to that of Case 2, one shows that A is either of the form $[a, +\infty)$ or $(a, +\infty)$, where a is the greatest lower bound of A.

Case 4. A has a lower bound and an upper bound. Let a be the greatest lower bound of A and let b be the least upper bound of A. Since A contains at least two distinct points, $a < b$. A point x, if it is to lie in A, must therefore lie in $[a, b]$, so that $A \subset [a, b]$. We claim that $a < x < b$ implies that $x \in A$. This implication follows from the fact that for any such point x, there must be points a' and b' with $a', b' \in A$ and $a \leqq a' < x < b' \leqq b$. Hence $(a, b) \subset A \subset [a, b]$. Consequently, A must be of one of the four forms (a, b), $[a, b)$, $(a, b]$, or $[a, b]$, depending on which, if any, of the two points a, b belongs to A.

We shall now prove that apart from the empty set and single points, the only connected subsets of the real line are intervals.

Theorem 3.4 A subset A of the real line that contains at least two distinct points is connected if and only if it is an interval.

Proof. We shall first show that if A is not an interval then it is not connected. If A is not an interval, then there are points a, b, c with $a < c < b$ and $a, b \in A$, whereas $c \notin A$. Let $P = (-\infty, c)$, $Q = (c, +\infty)$. P and Q are open subsets of the real line that satisfy the conditions of Lemma 2.3; hence A is not connected.

Conversely, we shall show that if A is not connected then A is not an interval. If A is not connected, by Lemma 2.4, there are closed subsets F and G of the real line such that $A \subset F \cup G$, $F \cap G \subset C(A)$ and both F and G contain a point of A. Assume that the notation is such that there is a point $a \in A \cap F$ and a point $b \in A \cap G$ with $a < b$. We shall find a point between a and b that is not in A. Let $G' = G \cap [a, b]$. Then G' is a closed non-empty subset of the real line and, consequently, contains its greatest lower bound c. We cannot have $a = c$, for then $A \cap F \cap G \neq \emptyset$, contradicting $F \cap G \subset C(A)$. Thus, $a < c$. Next, let $F' = F \cap [a, c]$. F' is also a closed non-empty subset of the real line and therefore contains its least upper bound d. In the event that $c = d$

we have $c \in F \cap G$, hence $c \notin A$ and A is not an interval. Otherwise $d < c$ and $(d, c) \cap (F \cup G) = \emptyset$, so that $(d, c) \cap A = \emptyset$, and again A does not contain a point between a and b and is therefore not an interval.

Exercises

1. Let $f: R \to R$ be continuous. Prove that the image under f of each interval is either a single point or an interval.
2. Prove that Euclidean n-space, R^n, and the standard n-cube, I^n, are connected for each value of n.
3. Prove that a homeomorphism $f: [a, b] \to [a, b]$ carries end points into end points.
4. Let X be a topological space. Prove that X is connected if and only if each open subset U of X different from $\emptyset$ and X has at least one boundary point.

4 Some Applications of Connectedness

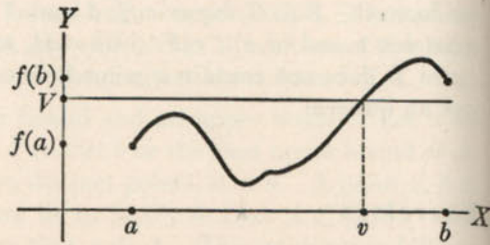
Theorem 4.1 (Intermediate-Value Theorem.) Let $f: [a, b] \to R$ be continuous and let $f(a) \neq f(b)$. Then for each number V between $f(a)$ and $f(b)$ there is a point $v \in [a, b]$ such that $f(v) = V$.

Proof. $[a, b]$ is connected, hence $f([a, b])$ is connected and is therefore an interval. Now, $f(a), f(b) \in f([a, b])$. Thus if V is between $f(a)$ and $f(b)$, since $f([a, b])$ is an interval, $V \in f([a, b])$; that is, there is a $v \in [a, b]$ such that $f(v) = V$.

Theorem 4.1 states that for each V between $f(a)$ and $f(b)$, the horizontal line $y = V$ intersects the graph of $y = f(x)$ at some point (v, V) with $a < v < b$, as indicated in Figure 13.

If the domain of a continuous real-valued function contains an interval $[a, b]$, then its restriction to $[a, b]$ is continu-

Figure 13



ous and we can assert that f must assume at least once each value between $f(a)$ and $f(b)$ over the interval $[a, b]$.

As a special case of the intermediate-value theorem, namely $V = 0$, we have,

Corollary 4.2 Let $f:[a, b] \to R$ be continuous. If $f(a)f(b) < 0$, then there is an $x \in [a, b]$ such that $f(x) = 0$.

Corollary 4.3 (Fixed-Point Theorem). Let $f:[0, 1] \to [0, 1]$ be continuous. Then there is a $z \in [0, 1]$ such that $f(z) = z$.

Proof. In the event that $f(0) = 0$ or $f(1) = 1$, the theorem is certainly true. Thus, it suffices to consider the case in which $f(0) > 0$ and $f(1) < 1$. Let $g:[0, 1] \to R$ be defined by

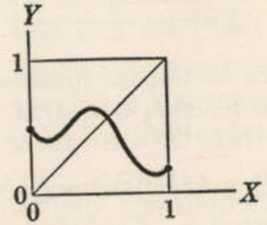
$$g(x) = x - f(x),$$

(therefore, if $g(z) = 0$, $f(z) = z$). g is continuous and $g(0) = -f(0) < 0$, whereas $g(1) = 1 - f(1) > 0$. Consequently, by Corollary 4.2, there is a $z \in [0, 1]$ such that $g(z) = 0$, whence $f(z) = z$.

We may interpret this theorem geometrically. Since $f:[0, 1] \to [0, 1]$, the graph of $y = f(x)$ is contained in the unit square defined by $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$. The point $(z, f(z))$ given by the theorem lies on both the graph of $y = f(x)$ and the line $y = x$. Hence the theorem asserts that the graph of $y = f(x)$ intersects the line $y = x$ in this square (see Figure 14), or equivalently, that in order for the curve which constitutes the graph to connect a point on the left-hand

edge of the square with a point on the right-hand edge of the square, the curve must intersect the diagonal of the square pictured in Figure 14. Although statements of this type seem

Figure 14



intuitively true, they are oftentimes most difficult to prove.

The reason for calling Theorem 4.3 a fixed-point theorem is that, if we think of $f:[0, 1] \to [0, 1]$ as a transformation that carries each point x of $[0, 1]$ into the point $f(x)$ of $[0, 1]$, then to say that $f(z) = z$ is to say that the transformation f leaves z "fixed."

There are many so-called "fixed-point" theorems, of which 4.3 is undoubtedly the simplest. In general, a fixed-point theorem is one that states that for a specified topological space X each continuous function $f:X \to X$ possesses a fixed point; that is, there is necessarily a $z \in X$ such that $f(z) = z$. One of the convenient facts about a fixed-point theorem is that if X and Y are homeomorphic topological spaces and a fixed-point theorem is true for X, then it is also true for Y.

Theorem 4.4 Let X and Y be homeomorphic topological spaces. Then each continuous function $h:X \to X$ possesses a fixed point if and only if each continuous function $k:Y \to Y$ possesses a fixed point.

Proof. Let $f:X \to Y$ and $g:Y \to X$ be a pair of continuous inverse functions. Let $k:Y \to Y$ be a continuous function so that we have the diagram

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & Y \arrow[d, "k"] \\ X & Y \arrow[l, "g"] \end{tikzcd}$$

and suppose that each continuous function $h: X \to X$ possesses a fixed point. Then the function $h = gkf: X \to X$ is continuous and there is a $z \in X$ such that $h(z) = z$. Let $w = f(z)$. We have

$$k(w) = k(f(z)) = fg(k(f(z))) = f(h(z)) = f(z) = w.$$

Thus, w is a fixed point of k. Since the hypotheses are symmetric with regard to X and Y, it also follows that if each continuous function $k: Y \to Y$ has a fixed point then so does each continuous function $h: X \to X$.

Any two closed intervals are homeomorphic. Since a fixed-point theorem holds for $[0, 1]$, we obtain

Corollary 4.5 Let $f: [a, b] \to [a, b]$ be continuous. Then there is a $z \in [a, b]$ such that $f(z) = z$.

Theorem 4.3 is a special case of the "Brouwer Fixed-Point Theorem," which we shall now state. Recall that in R^n, the unit n-cube I^n is defined as the set of points $(x_1, x_2, \ldots, x_n)$ whose coordinates satisfy the inequalities

$$0 \leqq x_i \leqq 1,$$

for $i = 1, 2, \ldots, n$.

Theorem 4.6 (Brouwer Fixed-Point Theorem.) Let $f: I^n \to I^n$ be continuous. Then there is a point $z \in I^n$ such that $f(z) = z$.

For $n = 1$, $I^1 = [0, 1]$, and Theorem 4.6 reduces to 4.3. We shall not prove this theorem. However, one can supply a

very suggestive argument for the truth of the theorem in the case $n = 2$. To this end we may, on the basis of Theorem 4.5, work with a topological space homeomorphic to I^2. If we think of I^2 as being a surface constructed of elastic material, we may conceive of a deformation or stretching by which we obtain a surface that is a disc; that is, the set of points (x_1, x_2) in the plane whose coordinates satisfy the inequality $x_1^2 + x_2^2 \leqq 1$. Thus, the disc is homeomorphic with I^2, and we may argue the validity of the fixed-point theorem with regard to the disc.

Let g be a continuous transformation of this disc into itself. Suppose that it were possible that for each point x of the disc, we had $g(x) \neq x$. Then for each point x in the disc, there would be a unique half-line L_x emanating from $g(x)$ and passing through x (see Figure 15). The half-line L_x

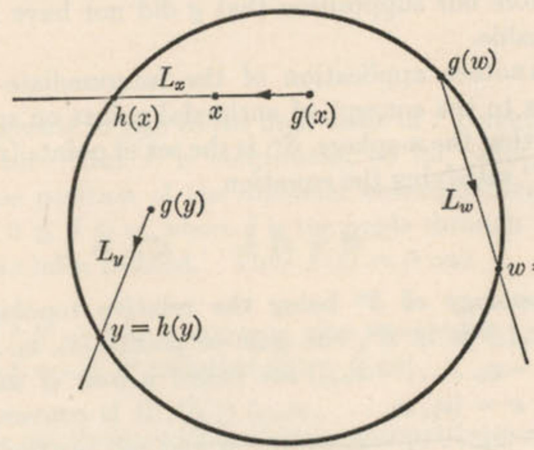


Figure 15

will contain a point on the boundary of the disc other than $g(x)$. Let us call this point $h(x)$. In particular, if y is a boundary point of the disc, then $h(y) = y$. This is true even if $g(y)$ itself is a boundary point, as may be seen by examining the various cases depicted in Figure 15. Using the given

transformation g we have thus constructed a new transformation h, which has the property that it carries each point of the disc into a boundary point and leaves each boundary point fixed (h is called a "retraction" since it retracts or pulls the interior of the disc onto its boundary while leaving the boundary fixed.)

We next argue that the transformation h is continuous, for the image $h(x)$ will vary by a small amount if we suitably restrict the variation of x. Though it is by no means simple to prove that no continuous transformation such as h can exist, our intuition should tell us that none can. For if there were a function such as h we should be able to retract the head of a drum onto the rim, although intuitively we know that we can do so only by ripping the drum head someplace; that is, by introducing a discontinuity. Since there is no function such as the retraction h, we have obtained a contradiction, and therefore our supposition that g did not have a fixed point is untenable.

Another application of the intermediate-value theorem relates to the concept of antipodal points on spheres. Let us recall that the n-sphere, S^n, is the set of points $(x_1, x_2, \ldots, x_{n+1})$ in R^{n+1} satisfying the equation

$$x_1^2 + x_2^2 + \ldots x_{n+1}^2 = 1,$$

the topology of S^n being the relative topology. If $(x_1, x_2, \ldots, x_{n+1})$ is in S^n, the pair of points $(x_1, x_2, \ldots, x_{n+1})$ and $(-x_1, -x_2, \ldots, -x_{n+1})$ are called a *pair of antipodal points*. Given $x = (x_1, x_2, \ldots, x_{n+1}) \in S^n$, it is convenient to denote $(-x_1, -x_2, \ldots, -x_{n+1})$ by $-x$ and call $-x$ the *antipodal point* of x. A pair $x, -x$ of antipodal points is the pair of end points of a diameter of the sphere. We shall be particularly interested in the 1-sphere, S^1, which is a circle.

Consider a continuous function $f:S^1 \to R$. If we define $F:S^1 \to R$ by $F(x) = f(x) - f(-x)$ for $x \in S^1$, we can show that $F(z) = 0$ for some $z \in S^1$; that is, $f(z) = f(-z)$, or f has

the same value at one or more pairs of antipodal points. The proof of this fact is motivated by the consideration that a value of F is determined by a diameter of the circle and a designation of one of its extremities as x and the other as $-x$. If we rotate this diameter through π radians, as indicated in Figure 16, then the initial value of F corresponding to

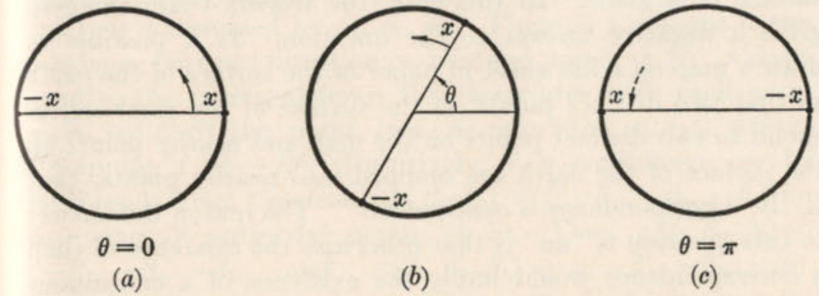


Figure 16

Figure 16a is opposite in sign to the final value of F corresponding to Figure 16c. But F is continuous, so its value must be zero for some position of the diameter corresponding to a value of θ with $0 \leqq \theta \leqq \pi$, where θ is the angle through which the diameter has been rotated. Thus $F(z) = 0$ and

Theorem 4.7 Let $f:S^1 \to R$ be continuous, then there exists a pair of antipodal points $z, -z \in S^1$ such that $f(z) = f(-z)$.

Whether or not the preceding discussion constitutes a proof of this theorem is legitimately a matter of opinion. Although it is hoped that it is convincing, a more analytic approach, which may appeal to many as being more rigorous, is indicated in the exercises.

One of the more significant generalizations of this theorem is called the Borsuk-Ulam theorem.

Theorem 4.8 (Borsuk-Ulam.) Let $f: S^n \to R^n$ be continuous, then there exists a pair of antipodal points $z, -z \in S^n$ such that $f(z) = f(-z)$.

We shall not prove this theorem. Theorem 4.7 is, of course, the case $n = 1$. The case $n = 2$ has some practical consequences. The 2-sphere, S^2, may be thought of as the surface of a globe. In this case, the Borsuk-Ulam theorem gives a negative answer to the question, "Is it possible to draw a map on a flat sheet of paper of the surface of the earth so that two distinct points on the surface of the earth correspond to two distinct points on the map and nearby points on the surface of the earth are mapped into nearby points, that is, the correspondence is continuous?" The reason the answer to this question is "no" is that otherwise the existence of such a correspondence would imply the existence of a continuous function $f: S^2 \to R^2$ that was one-one, and this possibility is ruled out by the Borsuk-Ulam theorem.

The case $n = 2$ also provides a solution to a problem often referred to as the "Ham-Sandwich Problem." This result states that, given three finite volumes U, V, W in R^3, there must exist at least one plane in R^3 that simultaneously bisects all three volumes; thus, regardless of the distribution of ham and cheese in a sandwich, one may, with one slice, divide the sandwich into two parts containing equal amounts of bread, ham, and cheese.

Let $\alpha_1, \alpha_2, \alpha_3$ be a set of direction cosines (so that $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1$). For each real number p, the plane

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = p \tag{1}$$

divides the three volumes. For each set of direction cosines choose $p = p(\alpha_1, \alpha_2, \alpha_3)$ so that the plane (1) bisects the volume U. In the event that there is more than one such value of p for which the volume U is bisected, the values of p constitute a closed interval and we choose the mid-point of this interval. [Note that in this manner the two planes

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = p(\alpha_1, \alpha_2, \alpha_3) \tag{2}$$

and

$$(-\alpha_1)x_1 + (-\alpha_2)x_2 + (-\alpha_3)x_3 = p(-\alpha_1, -\alpha_2, -\alpha_3)$$

coincide.] Let $v = v(\alpha_1, \alpha_2, \alpha_3)$ and $w = w(\alpha_1, \alpha_2, \alpha_3)$ be the measure of the volume of V and W, respectively, which lie on the same side of the plane (2) as the normal to this plane in the direction determined by $\alpha_1, \alpha_2, \alpha_3$. There is a one-one correspondence between direction cosines and points of S^2. Consequently, the correspondence that associates with each point $(\alpha_1, \alpha_2, \alpha_3) \in S^2$ the point $(v(\alpha_1, \alpha_2, \alpha_3), w(\alpha_1, \alpha_2, \alpha_3))$ defines a mapping $f: S^2 \to R^2$. Intuitively, f is continuous, so by the Bursuk-Ulam theorem $f(\beta_1, \beta_2, \beta_3) = f(-\beta_1, -\beta_2, -\beta_3)$ for some pair of antipodal points of S^2. Thus, $v(\beta_1, \beta_2, \beta_3) = v(-\beta_1, -\beta_2, -\beta_3)$ and $w(\beta_1, \beta_2, \beta_3) = w(-\beta_1, -\beta_2, -\beta_3)$. Since the plane $\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 = p(\beta_1, \beta_2, \beta_3)$ is the same as the plane $(-\beta_1)x_1 + (-\beta_2)x_2 + (-\beta_3)x_3 = p(-\beta_1, -\beta_2, -\beta_3)$, this plane is such that equal measures of the volumes V and W lie on either side of it; that is, it simultaneously bisects all three volumes.

For the student interested in pursuing some of the topics treated in this section, the paper by Tucker, *Some topological properties of disk and sphere*, Proceedings Canadian Mathematical Congress, pp. 285–309, is particularly recommended. Proofs of the Brouwer Fixed-Point Theorem may be found in many of the standard texts, including Lefschetz, *Introduction to Topology*; Cairns, *Introductory Topology*; Young and Hocking, *Topology*, etc. The Borsuk-Ulam theorem also occurs in the book by Lefschetz.

Exercises

1. Prove that the mapping $p: R \to S^1$ defined by

$$p(t) = (\cos t, \sin t)$$

for $t \in R$ is continuous and that therefore, for each continuous

function $g: S^1 \to R$, there is a continuous function $\varphi: R \to R$ such that the diagram

$$\begin{tikzcd} R \arrow[d, "p"'] \arrow[dr, "\varphi"] & \\ S^1 \arrow[r, "g"] & R \end{tikzcd}$$

is commutative. Let $f: S^1 \to R$ be continuous and define $F: S^1 \to R$ by $F(x) = f(x) - f(-x)$ for $x \in S^1$. Prove that $(Fp)(t) = -(Fp)(t + \pi)$, and that therefore there is a $z \in [0, \pi]$ such that $(Fp)(z) = 0$. Then show that if $x = p(z)$, $f(x) = f(-x)$, thereby proving Theorem 4.7.

2. Prove that given n finite volumes $U_1, U_2, \ldots, U_n$ in R^n there is a hyperplane

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = p$$

which simultaneously bisects all n volumes.

3. Let $F: R^2 \to R$ be a real-valued function defined and continuous on the plane. For each continuous function $f: [a, b] \to R$ we may define a new continuous function $Kf: [a, b] \to R$ by setting $Kf(t) = \int_a^t F(x, f(x))\, dx$, $t \in [a, b]$. Thus, if S is the set of continuous real-valued functions defined on $[a, b]$, K defines a transformation of S into itself. Prove that an element $g \in S$ is a fixed point of K if and only if g satisfies the differential equation $g'(x) = F(x, g(x))$ with initial condition $g(a) = 0$.

5 Components and Local Connectedness

In any topological space X, each point $a \in X$ belongs to a maximal connected subset of X called the "component of a."

Theorem 5.1 Let X be a topological space. For each point $a \in X$ there is a non-empty subset $\mathrm{Cmp}(a)$, called the *component* of a, with the property that $\mathrm{Cmp}(a)$ is connected and if D is any connected subset of X containing a, then $D \subset \mathrm{Cmp}(a)$.

Proof. There are connected subsets of X containing a for $\{a\}$ is such a subset. Let I be an indexing set for the family of connected subsets $(D_\alpha)_{\alpha \in I}$ containing a. We set $\mathrm{Cmp}(a) = \bigcup_{\alpha \in I} D_\alpha$. Thus, if D is any connected subset of X containing a, $D = D_\beta$ for some $\beta \in I$, whence $D \subset \mathrm{Cmp}(a)$. It remains to prove that $\mathrm{Cmp}(a)$ is connected. Suppose it is not. Then there are non-empty relatively open subsets A and B of $\mathrm{Cmp}(a)$ such that $A \cap B = \emptyset$ and $A \cup B = \mathrm{Cmp}(a)$. Assume the notation is such that $a \in A$ and let b be a point of B. Since $b \in \mathrm{Cmp}(a)$, $b \in D_\gamma$ for some connected subset D_γ of X containing a. Now $D_\gamma \subset \mathrm{Cmp}(a)$, hence $A' = A \cap D_\gamma$ and $B' = B \cap D_\gamma$ are non-empty relatively open subsets of D_γ. Furthermore, $A' \cap B' \subset A \cap B = \emptyset$ and $A' \cup B' = D_\gamma \cap (A \cup B) = D_\gamma$. Consequently, the supposition that $\mathrm{Cmp}(a)$ is not connected yields the contradiction that D_γ is not connected. Therefore $\mathrm{Cmp}(a)$ is connected.

Lemma 5.2 In a topological space X, let $b \in \mathrm{Cmp}(a)$. Then $\mathrm{Cmp}(b) = \mathrm{Cmp}(a)$.

Proof. Since $b \in \mathrm{Cmp}(a)$ and $\mathrm{Cmp}(a)$ is a connected set containing b, by Theorem 5.1, $\mathrm{Cmp}(a) \subset \mathrm{Cmp}(b)$. But $a \in \mathrm{Cmp}(a)$, hence $a \in \mathrm{Cmp}(b)$, so, by the same reasoning it follows that $\mathrm{Cmp}(b) \subset \mathrm{Cmp}(a)$ and therefore $\mathrm{Cmp}(a) = \mathrm{Cmp}(b)$.

Corollary 5.3 In a topological space X, for any two points $a, b \in X$, either $\mathrm{Cmp}(a) = \mathrm{Cmp}(b)$ or $\mathrm{Cmp}(a) \cap \mathrm{Cmp}(b) = \emptyset$.

Proof. If $\mathrm{Cmp}(a) \cap \mathrm{Cmp}(b) \neq \emptyset$, then for a point $c \in \mathrm{Cmp}(a) \cap \mathrm{Cmp}(b)$ we have, by Lemma 5.2, $\mathrm{Cmp}(c) = \mathrm{Cmp}(a)$ and $\mathrm{Cmp}(c) = \mathrm{Cmp}(b)$.

A subset of X that is the component of some point $a \in X$ is called a *component* of X. By Lemma 5.2, a component is the component of each of its points. By Corollary 5.3, two

components either coincide or are disjoint. The components of a topological space X thus constitute a particular example of what is known as a "partition" of a set.

Definition 5.4 Let X be a set and $(P_\alpha)_{\alpha\in I}$ an indexed family of non-empty subsets of X. $(P_\alpha)_{\alpha\in I}$ is called a *partition* of X if:

(i) $X = \bigcup_{\alpha\in I} P_\alpha$;
(ii) If $\alpha, \beta \in I$, $\alpha \neq \beta$, then $P_\alpha \cap P_\beta = \emptyset$.

In summary, we may state that the components of a topological space constitute a partition of the space into maximal connected subsets.

Theorem 5.5 Let $(Q_\alpha)_{\alpha\in I}$ be the indexed family of components of a topological space X. Then $(Q_\alpha)_{\alpha\in I}$ is a partition of X with the property that if D is any non-empty connected subset of X, then $D \subset Q_\beta$ for some $\beta \in I$.

Proof. Since $a \in \mathrm{Cmp}(a)$, each component is non-empty and each point of X is in some component. Thus $X = \bigcup_{\alpha\in I} Q_\alpha$. Finally, if D is a non-empty connected subset of X, then for some point $a \in D$, $D \subset \mathrm{Cmp}(a) = Q_\beta$ for some $\beta \in I$.

The following theorem implies that a component is necessarily a closed set.

Theorem 5.6 Let A be a connected subset of a topological space X and let $A \subset B \subset \overline{A}$. Then B is also connected.

Proof. We shall show that if B is not connected then A is not connected. For suppose there are open subsets P, Q of X such that $P \cap Q \subset C(B)$, $B \subset P \cup Q$, $P \cap B \neq \emptyset$, and $Q \cap B \neq \emptyset$. It would follow that $A \subset P \cup Q$ and since $C(B) \subset C(A)$, $P \cap Q \subset C(A)$. To prove that A is not connected we must show that $P \cap A \neq \emptyset$ and $Q \cap A \neq \emptyset$. If $P \cap A = \emptyset$, then A would be contained in the closed set $C(P)$, hence $\overline{A} \subset C(P)$ or $P \cap \overline{A} = \emptyset$. But this last relation would imply that $P \cap B = \emptyset$. Thus, $P \cap A \neq \emptyset$. Similarly, $Q \cap A \neq \emptyset$.

Corollary 5.7 The closure of a connected set is connected.

Corollary 5.8 In a topological space, each component is a closed set.

Proof. Let A be a component, say $A = \mathrm{Cmp}(a)$. Then $\overline{A}$ is a connected set containing a and therefore $\overline{A} \subset \mathrm{Cmp}(a) = A$. But $A \subset \overline{A}$, hence in this case $A = \overline{A}$ and A is closed.

It might be thought that a component must also be an open set, but it need not be as the following example will indicate. Let X be the subspace of the real line consisting of the points 0 and all numbers of the form $\frac{1}{n}$, n a positive integer. The only connected set containing 0 is $\{0\}$, thus $\mathrm{Cmp}(0) = \{0\}$. On the other hand $\{0\}$ is not a neighborhood of 0 and hence $\{0\}$ is not an open subset of X.

A sufficient condition for the components in a space to be open is that the space be "locally connected."

Definition 5.9 A topological space X is said to be *locally connected at a point* $a \in X$ if each neighborhood N of a contains a connected neighborhood U of a. A topological space X is said to be *locally connected* if it is locally connected at each of its points.

Lemma 5.10 Let X be a locally connected topological space and let Q be a component. Then Q is an open set.

Proof. Let $a \in Q$. Since X is locally connected there is at least one connected neighborhood N of a. But $Q = \mathrm{Cmp}(a)$, hence by Theorem 5.1, $N \subset Q$, which, in turn, implies that Q is a neighborhood of a. Thus, Q is a neighborhood of each of its points and therefore Q is open.

If X is locally connected at a then there are "arbitrarily small" connected neighborhoods of a, for, given any neighborhood N of a, there is a connected neighborhood $U \subset N$ that is at least as "small" as N. An equivalent formulation of local connectedness is obtained by utilizing the concept of basis for the neighborhoods at a.

Lemma 5.11 A topological space is locally connected at a point $a \in X$ if and only if there is a basis for the neighborhoods at a composed of connected subsets of X.

Proof. First, suppose that X is locally connected at a and let U_a be the collection of connected neighborhoods of a. Since every neighborhood N of a contains an element of U_a, U_a is a basis for the neighborhoods at a. Conversely, if there is a basis U_a for the neighborhoods of a consisting of connected sets, each neighborhood N must contain an element of U_a and therefore X is locally connected at a.

Exercises

1. Prove that a non-empty connected subset of a topological space that is both open and closed is a component.
2. Let X be a topological space that has a finite number of components. Prove that each component of X is both open and closed.
3. Let A be a connected subset of a locally connected space X. Then the subspace A is locally connected.
4. Let X and Y be homeomorphic topological spaces. Prove that any homeomorphism $f: X \to Y$ establishes a one-one correspondence between the components of X and the components of Y.
5. Prove that the product of two locally connected topological spaces is locally connected.
6. Prove that Euclidean n-space R^n and the standard n-cube I^n are locally connected.

6 Arcwise Connected Topological Spaces

In the three-dimensional geometry of the calculus, one often discusses a curve in terms of a parametric representation, usually written

$$x = f(t),$$
$$y = g(t),$$
$$z = h(t).$$

If not stated explicitly, it is generally understood that the three functions f, g, h are at least continuous, if not differentiable over some common interval $[a, b]$ as their domain, and therefore $F(t) = (f(t), g(t), h(t))$ defines a continuous function $F:[a, b] \to R^3$. The curve in question is, from this viewpoint, the image of $[a, b]$ under F; that is, $F([a, b])$. We may think of this curve as "connecting" the two points $F(a) = (f(a), g(a), h(a))$ and $F(b) = (f(b), g(b), h(b))$. Given two points $A, B \in R^3$, the question of whether or not there is a curve "connecting" A and B is therefore seen to be the same as the question of whether or not there is a continuous function $F:[a, b] \to R^3$ such that $F(a) = A$ and $F(b) = B$. Furthermore, the interval $[a, b]$ may just as well be restricted to $[0, 1]$, for using any homeomorphism $\varphi:[0, 1] \to [a, b]$, one may show that the required $F:[a, b] \to R^3$ exists if and only if a corresponding $G = F\varphi:[0, 1] \to R^3$ exists. These observations motivate the following two definitions:

Definition 6.1 Let X be a topological space. A continuous function $f:[0, 1] \to X$ is called an *arc* or *path* in X. The path f is said to *connect* or *join* the point $f(0)$ to the point $f(1)$. $f(0)$ is called the *initial point* and $f(1)$ is called the *terminal point* of the path f.

If f is a path in X, $f([0, 1])$ is called a *curve* in X.

Definition 6.2 A topological space X is said to be *arcwise connected* if, for each pair of points $u, v \in X$, there is a path f connecting u to v.

A non-empty subset A of a topological space X is said to be *arcwise connected* if the topological space A in the relative topology is arcwise connected.

The real line R is an arcwise connected space, for if a, b are two real numbers, the path $f:[0, 1] \to R$ defined by

$$f(t) = a + (b - a)t$$

for $t \in [0, 1]$ connects $f(0) = a$ and $f(1) = b$. R^n is also arcwise connected. This may be seen by either joining a given pair x, y of points of R^n by a path, or by using the general result that if X and Y are arcwise connected spaces, then so is $X \times Y$ (see Exercise 5 of this section). Another significant class of arcwise connected spaces are the spheres, S^n, for $n > 0$.

A path f in a topological space X whose initial and terminal points coincide is called a *closed* path or a *loop* in X. Though such paths play a significant role in topology, we shall not be concerned with them in this section.

If f is a path in a topological space X and g is a continuous mapping of X into a second topological space Y, then the composite function $gf:[0, 1] \to Y$ is a path in Y. This observation provides the basis for the proof of the result:

Theorem 6.3 Let Y be a topological space. If there exists an arcwise connected topological space X and a continuous mapping $g:X \to Y$, which is onto, then Y is arcwise connected.

Proof. Let $a, b \in Y$. Since $g:X \to Y$ is onto, there are points a', $b' \in X$ such that $g(a') = a$, $g(b') = b$. Since X is arcwise connected, there is a path f in X joining a' to b' and, consequently, the path gf joins a to b.

Note the necessity of the requirement that $g:X \to Y$ be onto. It follows that given homeomorphic topological spaces X and Y, X is arcwise connected if and only if Y is arcwise connected. Thus, arcwise connectedness is a topological property.

Arcwise connectedness is a stronger property than connectedness; that is, if a topological space X is arcwise connected then X is connected.

Theorem 6.4 Let X be an arcwise connected topological space, then X is connected.

Proof. Suppose X were not connected. Then there is a proper subset P of X which is both open and closed. Since P is proper, there is a point $a \in P$ and a point $b \in C(P)$. Let $f:[0, 1] \to X$ be a path from a to b. $f^{-1}(P)$ is a proper subset of $[0, 1]$ for $0 \in f^{-1}(P)$, $1 \notin f^{-1}(P)$. Since f is continuous, $f^{-1}(P)$ is both open and closed. But this contradicts the fact that $[0, 1]$ is connected. Therefore X is connected.

The converse of Theorem 6.4 is false. A counter-example to the converse, that is, a topological space that is connected but not arcwise connected, is the subspace of the plane consisting of the set of points (x, y) such that either

$$x = 0, \quad -1 \leqq y \leqq 1,$$

or

$$0 < x \leqq 1, \quad y = \cos \frac{\pi}{x}.$$

One may obtain some idea of this space by referring to Figure 17, where we have tried to show the main characteristics of this space. It is impossible to picture this space completely,

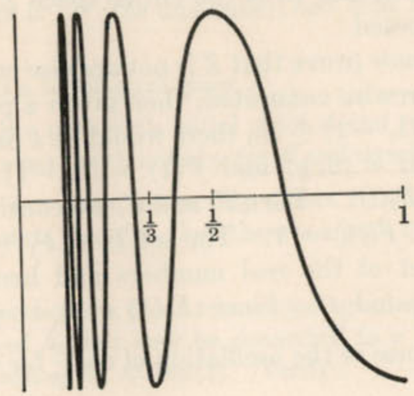


Figure 17

for, as the values of x approach 0, the oscillation of the graph $y = \cos \frac{\pi}{x}$ becomes more and more rapid.

It is not difficult to prove that this space is connected. First of all let us decompose this space into two subsets Z_1 and Z_2, where Z_1 is the set of points $(0, y)$, $-1 \leqq y \leqq 1$, on the Y-axis, and Z_2 is the complementary set consisting of those points (x, y), $0 < x \leqq 1$ and $y = \cos \frac{\pi}{x}$. The function $F(t) = \left(t, \cos \frac{\pi}{t}\right)$ defines a continuous mapping of the connected interval $(0, 1]$ onto Z_2, hence Z_2 is connected. To prove that the entire space $Z = Z_1 \cup Z_2$ is connected, we shall prove that $\bar{Z}_2 = Z$; that is, $Z_1 \subset \bar{Z}_2$. This is so because there are points of Z_2 arbitrarily close to each point of Z_1. For, let $(0, y) \in Z_1$ and let $\varepsilon > 0$ be given. We may find an even integer N sufficiently large so that $\frac{1}{N} < \varepsilon$. Now $\cos \frac{\pi}{1/N} = 1$ and $\cos \frac{\pi}{1/N + 1} = -1$, hence by the intermediate-value theorem there is a number $t \in \left[\frac{1}{N+1}, \frac{1}{N}\right]$ such that $\cos \frac{\pi}{t} = y$. The point $\left(t, \cos \frac{\pi}{t}\right)$ is in Z_1 and its distance from $(0, y)$ is less than ε. Thus $Z_1 \subset \bar{Z}_2$ and $\bar{Z}_2$ is the entire space Z. By Corollary 5.7, Z is connected.

We shall now prove that Z is not arcwise connected. Suppose Z were arcwise connected, then given a point $(0, y) \in Z_1$ and the point $(1, -1) \in Z_2$, there would be a path $F:[0,1] \to Z$ such that $F(0) = (0, y)$ and $F(1) = (1, -1)$. Let us write $F(t) = (F_1(t), F_2(t))$. Then F_1 and F_2 are continuous functions and $F_1(0) = 0$, $F_1(1) = 1$. The set $U = F_1^{-1}(\{0\})$ is a closed bounded subset of the real numbers and hence contains its least upper bound t^*. Since $F_1(1) \neq 0$, $t^* < 1$. We shall show that because of the oscillation of $\cos \frac{\pi}{x}$ for values of x close to zero, the function F_2 cannot be continuous at t^*. For each value of t such that $t^* < t \leqq 1$ we have $F_1(t) > 0$, hence $F(t) \in Z_2$ or $F(t) = \left(F_1(t), \cos \frac{\pi}{F_1(t)}\right)$; that is, $F_2(t) = \cos \frac{\pi}{F_1(t)}$. Let $\varepsilon \leqq 1$; we shall show that for each $\delta > 0$, $t^* + \delta \leqq 1$, there is a value of t such that $|t^* - t| < \delta$ whereas $|F_2(t^*) - F_2(t)| \geqq \varepsilon$. First $F_1(t^* + \delta) > 0$, hence there is an even integer N sufficiently large so that $F_1(t^*) = 0 < \frac{1}{N+1} < \frac{1}{N} < F_1(t^* + \delta)$. By the intermediate-value theorem, we may find $u, v \in [t^*, t^* + \delta]$ such that $F_1(u) = \frac{1}{N+1}$, $F_1(v) = \frac{1}{N}$. Since u, $v > t^*$ we have $F_2(u) = \cos \frac{\pi}{F_1(u)} = \cos (N + 1)\pi = -1$, $F_2(v) = \cos \frac{\pi}{F_1(v)} = \cos N\pi = 1$. Thus, if $F_2(t^*) \geqq 0$, $|F_2(t^*) - F_2(u)| \geqq 1 \geqq \varepsilon$, whereas if $F_2(t^*) \leqq 0$, $|F_2(t^*) - F_2(v)| \geqq 1 \geqq \varepsilon$, contradicting the continuity of F_2 at t^*. We have thus shown that no path such as F exists and that therefore our space Z is not arcwise connected.

Exercises

1. Prove directly by constructing appropriate paths that the topological spaces R^n, I^n (the unit cube), and $S^n (n > 0)$ are arcwise connected.

2. Verify that in a topological space X

(i) if there is a path with initial point A and terminal point B, then there is a path with initial point B and terminal point A, and

(ii) if there is a path connecting points A and B and a path connecting points B and C, then there is a path connecting points A and C.

3. The *arc component* of a point x in a topological space X is the set of all points of X that may be connected to x by a path in X. Denote this subset by ACmp(x). Verify:

(i) for each $x \in X$, $x \in \text{ACmp}(x)$;

(ii) for each $x, y \in X$, if $y \in \text{ACmp}(x)$, then $x \in \text{ACmp}(y)$;

(iii) for each $x, y, z \in X$, if $y \in \text{ACmp}(x)$ and $z \in \text{ACmp}(y)$, then $z \in \text{ACmp}(x)$;

(iv) for each $x \in X$, $\text{ACmp}(x)$ is arcwise connected;

(v) if A is an arcwise connected subset of X, then $A \subset \text{ACmp}(x)$ for some $x \in X$.

(vi) X is arcwise connected if and only if $X = \text{ACmp}(x)$ for some $x \in X$.

4. If A and B are arcwise connected subsets of a topological space X and $A \cap B \neq \emptyset$, then $A \cup B$ is arcwise connected.

5. If X and Y are arcwise connected topological spaces, then $X \times Y$ is arcwise connected. [*Hint:* Given $z = (x, y)$ and $z' = (x', y') \in X \times Y$, connect (x, y) to (x, y') and connect (x, y') to (x', y').] Thus, prove by induction that if X is arcwise connected, X^n is arcwise connected.

6. Prove that the space Z considered in this section is an example of a connected space that is not locally connected.

7 Homotopic Paths

The collection of points on and between the two concentric circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 2$ is called an *annulus*. It is easy to see that this annulus is arcwise connected. For example, given two points $p_0 = (x_0, y_0)$ and $p_1 = (x_1, y_1)$, one may construct a path from p_0 to p_1 by first traversing the radius on which p_0 lies until we reach a point whose distance from the origin is the same as that of p_1 and then traversing in a clockwise direction the circular arc from this point to p_1 (see Figure 18). Let us call this path F_0. Alternately, one may construct a second path F_1 from p_0 to p_1, by first traversing in a clockwise direction a circular arc from p_0 to

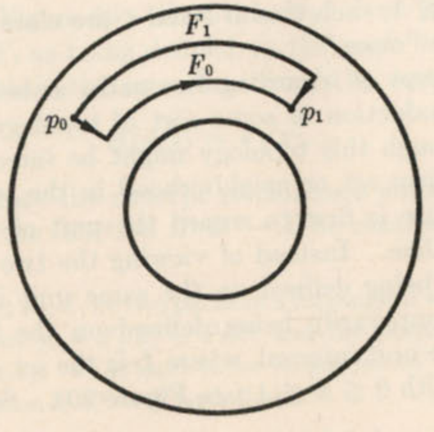


Figure 18

the radius on which p_1 lies and then traversing this radius until p_1 is reached. If, for the moment, we think of each of these two paths F_0 and F_1 as being represented by elastic strings with initial point p_0 and terminal point p_1, it is clear that in a given unit of time it would be possible to smoothly deform the path F_0 into the path F_1 (keeping p_0 and p_1 fixed throughout the deformation). This deformation might be

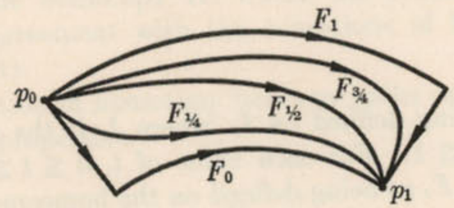


Figure 19

carried out so that at time $t = \frac{1}{4}$ the string lies over the curve $F_{1/4}$ of Figure 19, at time $t = \frac{1}{2}$ the string lies over $F_{1/2}$, and at time $t = \frac{3}{4}$ the string lies over $F_{3/4}$. We may thus conceive of the deformation of the path F_0 into the path F_1 as being accomplished by constructing an entire family of paths

F_t for $0 \leqq t \leqq 1$, such that if t and t' are close then the paths F_t and $F_{t'}$ are "close."

The concept of regarding two paths as being "close" implies the introduction of some sort of topology in this set of paths. Although this topology might be introduced directly by defining open set or neighborhood in the set of paths, an easier procedure is first to regard the unit of time as a unit interval on a line. Instead of viewing the two original paths F_0 and F_1 as being defined on the same unit interval, let us view F_0 as temporarily being defined on the homeomorphic image I_0 of the unit interval, where I_0 is the set of points $(x, 0)$ in the plane with $0 \leqq x \leqq 1$ (see Figure 20). Similarly, let us

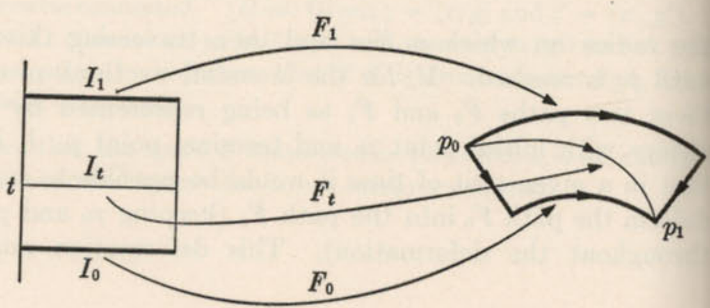


Figure 20

view F_1 as being defined on I_1, where I_1 is the set of points $(x, 1)$, $0 \leqq x \leqq 1$. For each value of t, $0 \leqq t \leqq 1$, we may view the path F_t as being defined on the homeomorphic image of the unit interval I_t, where I_t is the set of points (x, t), $0 \leqq x \leqq 1$. If we have such a situation, we may define a function $H: I^2 \to X$, where I^2 is the unit square and X is our annulus, by setting

$$H(x, t) = F_t(x, t),$$

as depicted in Figure 20. Equivalently, if we insist on viewing each path F_t as being defined on the same unit interval I, we may still obtain the same function H by setting

$$H(x, t) = F_t(x).$$

We now introduce the concept of closeness amongst paths by requiring that the function $H: I^2 \to X$ be continuous.

Definition 7.1 Let F_0, F_1 be two paths in a topological space X with the same initial point $p_0 = F_0(0) = F_1(0)$ and the same terminal point $p_1 = F_0(1) = F_1(1)$. F_0 is said to be *homotopic* to F_1 if there is a continuous function $H: I^2 \to X$ such that

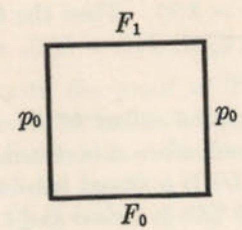
$$\begin{aligned} H(0, t) &= p_0, & 0 \leqq t \leqq 1, \\ H(1, t) &= p_1, & 0 \leqq t \leqq 1, \\ H(x, 0) &= F_0(x), & 0 \leqq x \leqq 1, \\ H(x, 1) &= F_1(x), & 0 \leqq x \leqq 1. \end{aligned}$$

The function H is called a homotopy from F_0 to F_1.

In this event we say that the path F_0 is deformable into the path F_1 with fixed end points. One may illustrate the fact that a path F_0 is homotopic to F_1 by indicating that I^2 is the domain of the homotopy H, where the boundary of I^2 is mapped in agreement with the conditions of Definition 7.1 (see Figure 21).

The relation of homotopy between paths satisfies the following three properties.

Figure 21



Theorem 7.2 Let F_0, F_1, F_2 be three paths in a topological space X with the same initial point p_0 and the same terminal point p_1.

(i) F_0 is homotopic to itself.

(ii) If F_0 is homotopic to F_1 then F_1 is homotopic to F_0.

(iii) If F_0 is homotopic to F_1 and F_1 is homotopic to F_2 then F_0 is homotopic to F_2.

Proof. To show that F_0 is homotopic to itself we need only define $H:I^2 \to X$ by $H(x, t) = F_0(x)$. Next, suppose that F_0 is homotopic to F_1 so that there is a homotopy $H:I^2 \to X$ from F_0 to F_1. For each $(x, t) \in I^2$, set $H'(x, t) = H(x, 1 - t)$. Then H' is easily seen to be a homotopy from F_1 to F_0. To prove (iii), first let G be a homotopy from F_0 to F_1 and let H be a homotopy from F_1 to F_2. We may construct a homotopy from F_0 to F_2 in stages. First, we alter H to a function H' defined for (x, t') with $1 \leq t' \leq 2$ so that G and H' together constitute a function K' defined for (x, t) with $0 \leq t \leq 2$. Finally, we compress K' to a function K again defined on I^2. The four diagrams of Figure 22 depict this process. To this end let $H'(x, t') = H(x, t' - 1), 0 \leq x \leq 1, 1 \leq t' \leq 2$. We then have two functions G and H', G defined on the subset $A = I^2$ of the plane and H' defined on the subset B consisting of the points (x, t) such that $0 \leq x \leq 1$ and $1 \leq t \leq 2$. The set $A \cap B$ consists of the points $(x, 1)$, $0 \leq x \leq 1$ and therefore we have $G(x, 1) = F_1(x)$, $H'(x, 1) = H(x, 0) = F_1(x)$; that is, G and H' agree in their common domain of definition. We shall now prove a lemma that asserts that together G and H' define a continuous function $K':A \cup B \to X$.

Lemma 7.3 Let A, B be closed subsets of a topological space Z. Let $g:A \to X$ and $h:B \to X$ be continuous functions with the property that for $z \in A \cap B$, $g(z) = h(z)$. Then the function $k:A \cup B \to X$ defined by $k(z) = g(z)$, $z \in A$, $k(z) = h(z)$, $z \in B$, is a continuous extension of g and h.

Proof. Let U be a closed subset of X. Then $g^{-1}(U)$ is a relatively closed subset of A and, since A is closed, $g^{-1}(U)$ is a closed subset of Z. Similarly, $h^{-1}(U)$ is a closed subset of Z. But $k^{-1}(U) = g^{-1}(U) \cup h^{-1}(U)$, hence $k^{-1}(U)$ is closed and k is continuous.

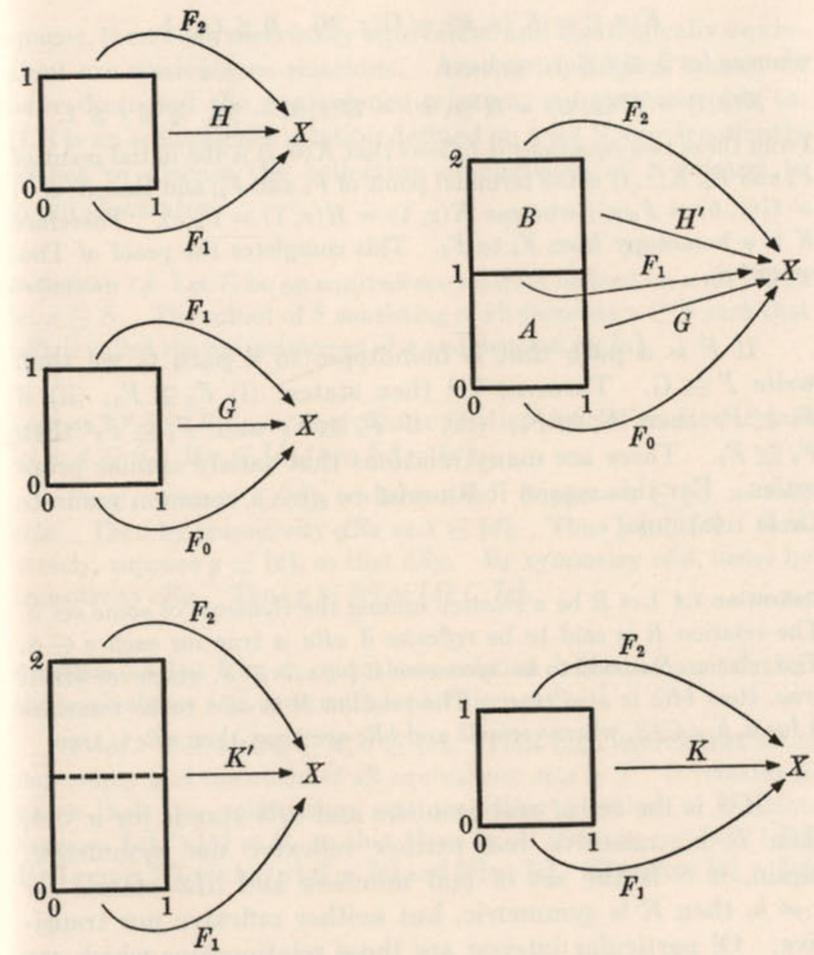


Figure 22

Continuing now with the proof of Theorem 7.2, the function $K':A \cup B \to X$ defined by $K'(x, t) = G(x, t)$, $(x, t) \in A$, $K'(x, t) = H'(x, t)$, $(x, t) \in B$ is continuous. We finally "compress" K' to the function $K:I^2 \to X$ defined by $K(x, t) = K'(x, 2t)$, $(x, t) \in I^2$. To recapitulate, for $(x, t) \in I^2$ with $0 \leq t \leq \frac{1}{2}$, we have

$$K(x, t) = K'(x, 2t) = G(x, 2t), \quad 0 \leqq t \leqq \tfrac{1}{2},$$

whereas for $\frac{1}{2} \leqq t \leqq 1$, we have

$$K(x, t) = K'(x, 2t) = H'(x, 2t) = H(x, 2t - 1), \quad \tfrac{1}{2} \leqq t \leqq 1.$$

From these two equations it follows that $K(0, t)$ is the initial point of F_0 and F_2, $K(1, t)$ is the terminal point of F_0 and F_2, and that $K(x, 0) = G(x, 0) = F_0(x)$, whereas $K(x, 1) = H(x, 1) = F_2(x)$. Therefore K is a homotopy from F_0 to F_2. This completes the proof of Theorem 7.2.

If F is a path that is homotopic to a path G we shall write $F \cong G$. Theorem 7.2 then states: (i) $F_0 \cong F_0$; (ii) if $F_0 \cong F_1$ then $F_1 \cong F_0$; (iii) if $F_0 \cong F_1$ and $F_1 \cong F_2$ then $F_0 \cong F_2$. There are many relations that satisfy similar properties. For this reason it is useful to give a common name to these relations.

Definition 7.4 Let R be a relation among the elements of some set S. The relation R is said to be *reflexive* if aRa is true for each $a \in S$. The relation R is said to be *symmetric* if for $a, b \in S$, whenever aRb is true, then bRa is also true. The relation R is said to be *transitive* if for $a, b, c \in S$, whenever aRb and bRc are true, then aRc is true.

If S is the set of real numbers and aRb stands for $a < b$, then R is transitive but neither reflexive nor symmetric. Again, if S is the set of real numbers and aRb stands for $a \neq b$, then R is symmetric, but neither reflexive nor transitive. Of particular interest are those relationships which are reflexive, symmetric, and transitive.

Definition 7.5 A relation R on a set S is called an *equivalence relation* on S if R is reflexive, symmetric, and transitive.

Theorem 7.2 thus states that being homotopic is an equivalence relation. There may be many equivalence relations defined on the same set. For example, if S is the set of metric

spaces, then both metrically equivalent and topologically equivalent are equivalence relations. Among topological spaces we have discussed the equivalence relation, is homeomorphic to. If R is an equivalence relation defined on a set S, one frequently wishes to discuss the collection of elements of S related to given element a.

Definition 7.6 Let R be an equivalence relation defined on a set S and let $a \in S$. The subset of S consisting of all elements $x \in S$ such that aRx is called the *equivalence set* of a and denoted by $[\![a]\!]$.

Lemma 7.7 Let R be an equivalence relation defined on a set S and let $c, d \in S$. If $c \in [\![d]\!]$ then $[\![c]\!] = [\![d]\!]$.

Proof. Since $c \in [\![d]\!]$, we have dRc. Suppose $x \in [\![c]\!]$, so that cRx. Then by transitivity dRx or $x \in [\![d]\!]$. Thus $[\![c]\!] \subset [\![d]\!]$. Conversely, suppose $y \in [\![d]\!]$, so that dRy. By symmetry cRd, hence by transitivity cRy. Thus $y \in [\![c]\!]$ or $[\![d]\!] \subset [\![c]\!]$.

Corollary 7.8 Let R be an equivalence relation defined on a set S; then the equivalence sets constitute a partition of S.

Proof. For each $a \in S$, $a \in [\![a]\!]$. Thus each equivalence set is non-empty and the union of all equivalence sets is S. It remains to prove that two equivalence sets are either identical or disjoint. Suppose $[\![a]\!] \cap [\![b]\!] \neq \emptyset$, so that there is an element $c \in [\![a]\!] \cap [\![b]\!]$. By Lemma 7.7, we have $[\![a]\!] = [\![c]\!]$ and $[\![b]\!] = [\![c]\!]$. Therefore $[\![a]\!] = [\![b]\!]$.

Exercises

1. Let F_0, F_1 be paths in a topological space X such that the terminal point of F_0 is the initial point of F_1. Define a function $F_0 \cdot F_1$ by setting

$$F_0 \cdot F_1(t) = F_0(2t), \qquad 0 \leqq t \leqq \tfrac{1}{2},$$
$$F_0 \cdot F_1(t) = F_1(2t - 1), \quad \tfrac{1}{2} \leqq t \leqq 1.$$

Prove that $F_0 \cdot F_1$ is a path in X whose initial point is the initial point of F_0 and whose terminal point is the terminal point of F_1.

Let F_2 be another path whose initial point is the terminal point of F_1. Verify that $(F_0 \cdot F_1) \cdot F_2 \neq F_0 \cdot (F_1 \cdot F_2)$, but $(F_0 \cdot F_1) \cdot F_2 \cong F_0 \cdot (F_1 \cdot F_2)$.

2. With F_0 and F_1 as in Exercise 1, let $F_0 \cong G_0$ and $F_1 \cong G_1$. Prove that $F_0 \cdot F_1 \cong G_0 \cdot G_1$.
3. Let X be a topological space and $x \in X$. The "constant" path at x, e_x, is the path defined by

$$e_x(t) = x, \quad 0 \leqq t \leqq 1.$$

Let F be any path in X with initial point p_0 and terminal point p_1. Prove that $e_{p_0} \cdot F \cong F$ and $F \cdot e_{p_1} \cong F$.

4. Let F be a path in a topological space X. Define a path F^{-1} by setting

$$F^{-1}(t) = F(1 - t), \quad 0 \leqq t \leqq 1.$$

Prove that $F \cdot F^{-1} \cong e_{p_0}$ if e_{p_0} is the constant path at the initial point p_0 of F and that $F^{-1} \cdot F \cong e_{p_1}$ if e_{p_1} is the constant path at the terminal point of F.

5. Let Z be the set of integers. For a non-zero integer n, we say that two integers a, b are *congruent modulo n* if $a - b$ is divisible by n. Prove that congruence modulo n is an equivalence relation. Describe the equivalence sets and determine how many there are.
6. Let A, B be subsets of a topological space X and let $f: A \to Z$, $g: B \to Z$ be continuous. Verify that Lemma 7.3 is equivalent to the statement that if A, B are closed and the diagram

$$\begin{tikzcd}
& A \arrow[dr, "f"] & \\
A \cap B \arrow[ur, "i"] \arrow[dr, "i"'] & & Z \\
& B \arrow[ur, "g"'] &
\end{tikzcd}$$

is commutative, where i stands for the various inclusion maps, then there is a commutative diagram

$$\begin{tikzcd}
& A \arrow[dr, "i"] \arrow[drr, "f"] & & \\
A \cap B \arrow[ur, "i"] \arrow[dr, "i"'] & & A \cup B \arrow[r, "k"] & Z \\
& B \arrow[ur, "i"'] \arrow[urr, "g"'] & &
\end{tikzcd}$$

where k is also continuous. Prove that the same conclusion holds if A and B are open.

8 Simple Connectedness

In discussing homotopic paths, an equivalence set of homotopic paths is called a *homotopy class.* Of particular interest are the homotopy classes of closed paths; that is, paths with the same initial and terminal point. Among these homotopy classes there is, for each point z in a topological space, the homotopy class $[\![e_z]\!]$, where e_z is the *constant* path at z defined by $e_z(s) = z$, $0 \leqq s \leqq 1$.

If f is a closed path at a point z of a topological space Z, and $f \in [\![e_z]\!]$, then a homotopy from f to e_z satisfies the conditions

$$\begin{aligned}
H(x, 0) &= f(x), & 0 &\leqq x \leqq 1, \\
H(x, 1) &= z, & 0 &\leqq x \leqq 1, \\
H(0, t) &= z, & 0 &\leqq t \leqq 1, \\
H(1, t) &= z, & 0 &\leqq t \leqq 1.
\end{aligned}$$

Pictorially, $H: I^2 \to Z$ carries the unit square into Z in such a manner that along the lower boundary of I^2, H agrees with f, whereas all other points on the boundary of I^2 are carried into z, as indicated in Figure 23. In effect, H deforms the closed

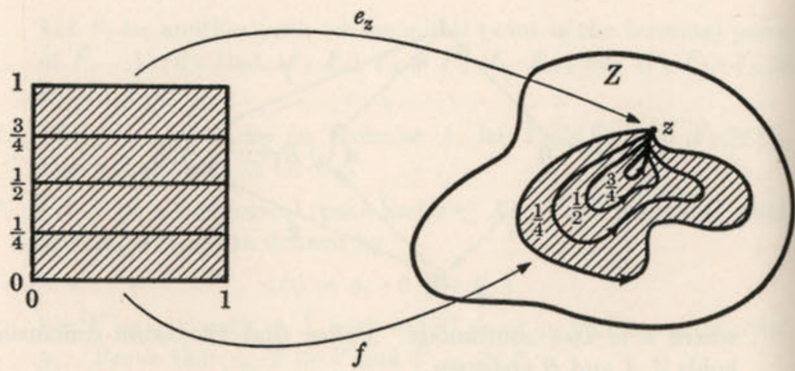


Figure 23

path f into the point z. Intuitively, the possibility of carrying out this deformation corresponds to the fact that the curve traced out by f does not enclose any holes in the space Z.

Definition 8.1 A topological space Z is said to be *simply connected* if at each point $z \in Z$ there is only one homotopy class of closed paths, or equivalently, if at each point $z \in Z$ any closed path at z is homotopic to the constant path at z.

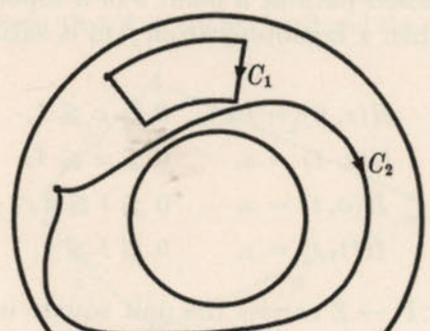


Figure 24

One can prove that an annulus is not simply connected, for, although a closed path such as c_1 (see Figure 24) is homotopic to a constant path, a closed path such as c_2 is not homotopic to a constant path. On the other hand, a disc is simply connected.

To determine whether or not an arcwise connected space is simply connected, it suffices to examine the closed paths at a single point.

Theorem 8.2 Let Z be an arcwise connected topological space and let $z \in Z$. Z is simply connected if and only if there is exactly one homotopy class of closed paths at z.

The proof of Theorem 8.2 consists primarily of the extension of various functions in order to obtain the desired homotopy. Suppose we have two closed paths g and h at a point $y \in Z$ and we know there is only one homotopy class of closed paths at the point z. Since Z is arcwise connected there is a path f from z to y. Using g and h, we can construct two closed paths g_f and h_f at z, by spending one-third of the unit interval tracing out f, the second third of the unit interval tracing out g and h respectively, and the final third of the unit interval returning to z along the reversal of f.

Definition 8.3 Let $f: I \to Z$ be a path in a topological space Z. By $f^{-1}: I \to Z$ is meant the path defined by

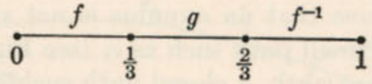
$$f^{-1}(t) = f(1 - t), \quad t \in I.$$

Let $z = f(0)$ and $y = f(1)$ be the initial and terminal points of f and let g be a closed path at y. By $g_f: I \to Z$ is meant the path defined by

$$\begin{aligned} g_f(t) &= f(3t), & 0 &\leqq t \leqq \tfrac{1}{3}, \\ g_f(t) &= g(3t - 1), & \tfrac{1}{3} &\leqq t \leqq \tfrac{2}{3}, \\ g_f(t) &= f(3 - 3t), & \tfrac{2}{3} &\leqq t \leqq 1. \end{aligned}$$

The path g_f may be viewed as arising in the fashion indicated by Figure 25, the appropriate paths being defined on

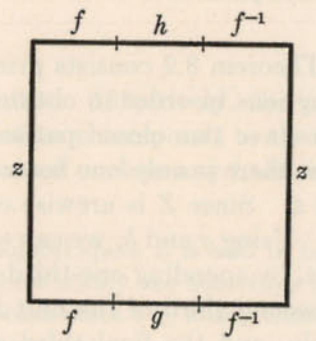
Figure 25



the indicated segments of the unit interval to obtain the closed path g_f.

Continuing our discussion of Theorem 8.2, g_f and h_f are closed paths at z. Since any two closed paths at z are homotopic, there is a homotopy K from g_f to h_f. We may picture K as being defined in accordance with Figure 26. Our aim is to

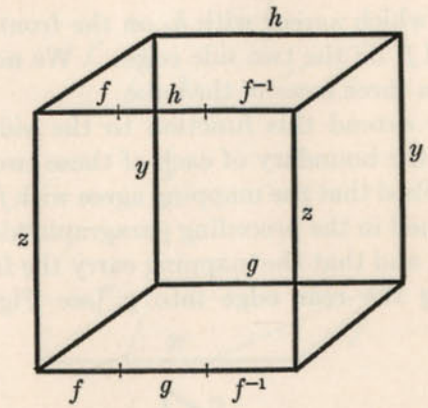
Figure 26



utilize this homotopy at z to obtain a homotopy L from g to h at y. In order to see how the homotopy L is constructed, let us view K as being defined on the front face of a cube and attempt to define L on the rear face. We shall accomplish this end by extending K to a mapping of the entire cube in such a way that the restriction to the rear face is the desired homotopy L. To achieve this result, the mapping on the rear face restricted to the lower edge must agree with g, restricted to the upper edge it must agree with h, and the vertical edges of the rear face must be carried into y (see Figure 27).

We begin the extension of K to the entire cube by first extending K to the four faces joining the front and rear of the cube, subject to the restrictions imposed in Figure 27.

Figure 27



First, let us examine how we may prescribe the mapping on the base of the cube. So far, on the front edge we have g_f and on the rear edge we have g (Figure 28a). Let P be the

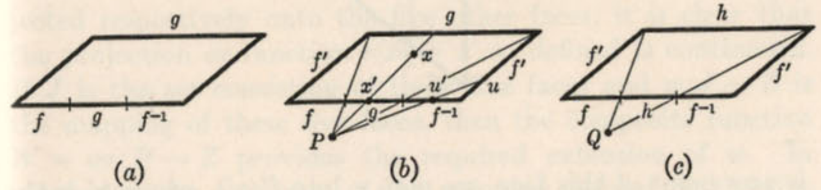


Figure 28

point in the plane of this face determined by the intersecting straight lines of Figure 28b. If x is any point on the base, we map x into $g_f(x')$, where x' is the projection of x from P on the front edge. There are two facts worth noting about this mapping of the base. First of all, this procedure will map the rear edge in agreement with g. Second, the restriction f' of this mapping to the two side edges maps these edges so that they trace out the same curve as f. A similar procedure results in a mapping defined on the upper face of the cube

(Figure 28c), which agrees with h_f on the front edge, h on the rear edge, and f' on the two side edges. We now have a function defined on three faces of the cube.

Next, we extend this function to the side faces of the cube. Along the boundary of each of these two faces we have already prescribed that the mapping agree with f', the modification of f obtained in the preceding paragraph, along the bottom and top edges, and that the mapping carry the front edge into z while carrying the rear edge into y (see Figure 29). If x

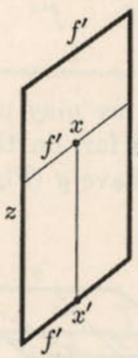
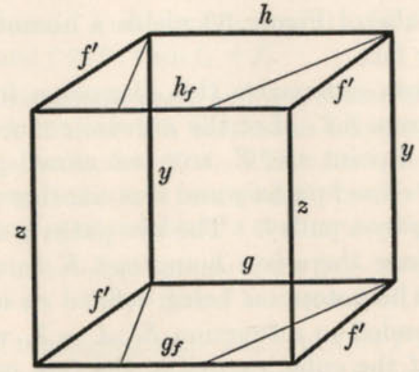


Figure 29

is any point of this face, we map x into $f'(x')$, where x' is the point of the bottom edge that x lies above. This definition of the mapping on the side faces is equivalent to prescribing that each horizontal line in the face be mapped in agreement with f'.

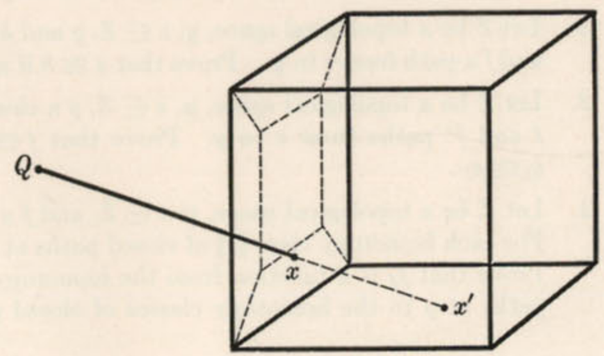
We have now extended our mapping to all the faces of the cube with the exception of the rear face, as indicated in Figure 30. In general, any mapping that is defined on all but one face of a cube may be extended to the entire cube. The procedure is to project each point x of the cube onto a point x' of one of the faces on which the mapping is already defined and let x be mapped into the same point that x' is mapped into. This procedure is illustrated in Figure 31,

Figure 30



where a mapping has been defined on all faces except the face on the left. Each point x of the cube is projected from a point Q a unit distance above the centroid to the left face onto a point x' of one of the remaining five faces. Though the analytic expression for this projection is complicated, since there are five distinct regions of the cube that are projected respectively onto the five other faces, it is clear that the projection or function $r(x) = x'$ so defined is continuous. If J is the set consisting of these five faces and $w: J \to Z$ is the mapping of these five faces, then the composite function $W = wr: I^3 \to Z$ provides the required extension of w. In particular, the extension of the mapping defined on the five

Figure 31



faces of the cube of Figure 30 yields a homotopy of g with h along the rear face.

Let us now summarize this discussion in the form of a proof of Theorem 8.2. Let the arcwise connected space Z be such that, at a point $z \in Z$, any two closed paths are homotopic. Given closed paths g and h at another point $y \in Z$, we may join z to y by a path f. The two paths g_f and h_f are closed paths at z, hence there is a homotopy K from g_f to h_f. We may view this homotopy as being defined on a face of a cube. K is then extended to a function $K': J \to Z$, where J consists of five faces of the cube, excluding the face opposite the face on which K is defined, and the restriction of K' to the boundary of this remaining face carries two opposite edges into y, whereas it maps the two remaining edges in agreement with g and h. K' is finally extended to a function $K'': I^3 \to Z$ and K'' restricted to this last face yields a homotopy of g and h. Thus Z is simply connected.

An Introduction to Algebraic Topology by Wallace is particularly recommended to those students who wish to investigate homotopy theory and related topics.

Exercises

1. Let X and Y be homeomorphic topological spaces. Prove that X is simply connected if and only if Y is simply connected.
2. Let Z be a topological space, $y, z \in Z$, g and h closed paths at y, and f a path from z to y. Prove that $g \cong h$ if and only if $g_f \cong h_f$.
3. Let Z be a topological space, $y, z \in Z$, g a closed path at y, and f and f' paths from z to y. Prove that $f \cong f'$ if and only if $g_f \cong g_{f'}$.
4. Let Z be a topological space, $y, z \in Z$, and f a path from z to y. For each homotopy class $[\![g]\!]$ of closed paths at y set $f_\#[\![g]\!] = [\![g_f]\!]$. Prove that $f_\#$ is a function from the homotopy classes of closed paths at y to the homotopy classes of closed paths at z that is one-one and onto. Furthermore, prove that if f' is a second path from z to y and $f \cong f'$, then $f_\# = f'_\#$.
5. Let X be a topological space and A a subset of X. A is called a *retract* of X if there is a continuous mapping $r: X \to A$ such that $r(a) = a$ for each $a \in A$. Prove that if A is a retract of X, then every continuous mapping $f: A \to Z$ of A into a topological space Z is extendable to a continuous mapping $F: X \to Z$.
6. Let J be the union of all but one of the $(n - 1)$-dimensional faces of the n-cube I^n. Prove that J is a retract of I^n.
7. Let X, Y be topological space and $f, g: X \to Y$ continuous functions. f is said to be *freely homotopic* to g if there is a continuous function $H: X \times I \to Y$ such that $H(x, 0) = f(x)$, $H(x, 1) = g(x)$ for each $x \in X$ [thus the family of functions $f_t: X \to Y$ defined by $f_t(x) = H(x, t)$ may be thought of as deforming the function f into the function g]. Prove that freely homotopic is an equivalence relation.
8. A topological space X is said to be *contractible to a point* $z \in X$ if there is a "homotopy" $H: X \times I \to X$ such that $H(x, 0) = x$, $H(x, 1) = z$ for each $x \in X$, and $H(z, t) = z$ for each $t \in I$. Prove that a space which is contractible to a point is simply connected.
9. Prove that for each positive integer n, R^n and I^n are contractible to a point.

Compactness

V

1 Introduction

A closed and "bounded" subset A of the real line R is characterized by the fact that for each collection $(O_\alpha)_{\alpha \in I}$ of open subsets of R such that $A \subset \cup_{\alpha \in I} O$, there is a finite subcollection $O_{\alpha_1}, O_{\alpha_2}, \ldots, O_{\alpha_n}$ with $A \subset \bigcup_{i=1}^{n} O_{\alpha_i}$. This second property is stated in terms that are applicable to any topological space. If this property holds in a particular topological space, the space is said to be "compact." The closed and bounded subsets of R^n are precisely the compact subspaces of R^n. This fact can be either proved directly or established by proving that the product of two compact spaces is itself compact. In metrizable spaces there is an alternate formulation of compactness; namely, that each infinite subset has a "point of accumulation."

Compactness, like connectedness and arcwise connectedness, is a "global" property, in that it depends on the nature of the entire space. The advantage in compact spaces is that one may study the whole space by studying a finite number of open subsets. We shall see this when we prove that a continuous function $f: X \to Y$ from a compact metric space X to a metric space Y is "uniformly continuous." In conclusion we shall examine some compact surfaces that may be formed by "identifying" edges of a rectangle.

2 Compact Topological Spaces

Definition 2.1 Let X be a set, B a subset of X, and $(A_\alpha)_{\alpha \in I}$ an indexed family of subsets of X. The collection $(A_\alpha)_{\alpha \in I}$ is called a *covering* of B or is said to *cover* B if

$$B \subset \cup_{\alpha \in I} A_\alpha.$$

If, in addition, the indexing set I is finite, $(A_\alpha)_{\alpha \in I}$ is called a *finite covering* of B.

In the event that $B = X$, $(A_\alpha)_{\alpha \in I}$ is a covering of X if $X = \cup_{\alpha \in I} A_\alpha$. Let X be a topological space and for each $x \in X$ let N_x be a neighborhood of x. Then $(N_x)_{x \in X}$ is a covering of X. For each integer n, let $A_n = [n, n+1]$. Then $(A_n)_{n \in Z}$, where Z is the set of integers, is a covering of the set R of real numbers. Similarly, if for each ordered pair (m, n) of integers we let $A_{m,n}$ be the set of points $(x_1, x_2) \in R^2$ such that $m \leqq x_1 \leqq m+1$, $n \leqq x_2 \leqq n+1$, then $(A_{m,n})_{(m,n) \in Z \times Z}$ is a covering of R^2. As a final example of a covering, let $X = R$ and let $B = (0, 1]$. If we set $A_1 = (\frac{1}{2}, 2)$, $A_2 = (\frac{1}{3}, 1)$, $A_3 = (\frac{1}{4}, \frac{1}{2})$, and in general, for each positive integer $n > 1$, set $A_n = \left(\frac{1}{n+1}, \frac{1}{n-1}\right)$, then $(A_n)_{n \in N}$, where N is the set of natural numbers, is a covering of B.

Definition 2.2 Let X be a set and let $(A_\alpha)_{\alpha\in I}$, $(B_\beta)_{\beta\in J}$ be two coverings of a subset C of X. If for each $\alpha \in I$, $A_\alpha = B_\beta$ for some $\beta \in J$, then the covering $(A_\alpha)_{\alpha\in I}$ is called a *subcovering* of the covering $(B_\beta)_{\beta\in J}$.

Thus $(A_\alpha)_{\alpha\in I}$ is a subcovering of $(B_\beta)_{\beta\in J}$ if "every A_α is a B_β." In particular, if $(B_\beta)_{\beta\in J}$ is a covering of a subset C, and I is a subset of J such that $(B_\beta)_{\beta\in I}$ is also a covering of C, then $(B_\beta)_{\beta\in I}$ is a subcovering of $(B_\beta)_{\beta\in J}$. Let Q be the set of rational numbers and for each $q \in Q$, set $B_q = [q, q+1]$. Then $(B_q)_{q\in Q}$ is a covering of the real numbers R. If again we let Z be the set of integers and $A_n = [n, n+1]$, then $(A_n)_{n\in Z}$ is a subcovering of $(B_q)_{q\in Q}$.

Suppose that $f:X \to Y$ is a continuous function from a topological space X to a metric space Y. Given $\varepsilon > 0$, the continuity of f gives rise to a covering of X in the following manner. For each $x \in X$, given this $\varepsilon < 0$, there is an open neighborhood U_x of x such that the images under f of all points of U_x are within ε of $f(x)$, or equivalently, $f(U_x) \subset S(f(x);\varepsilon)$. The family $(U_x)_{x\in X}$ of these subsets of X is clearly a covering of X. This covering has the additional property that it is composed of open sets. We shall, naturally, refer to such a covering as an "open" covering.

Definition 2.3 Let X be a topological space and B a subset of X. A covering $(A_\alpha)_{\alpha\in I}$ of B is said to be an *open covering* of B if for each $\alpha \in I$, A_α is an open subset of X.

Definition 2.4 A topological space X is said to be *compact* if for each open covering $(U_\alpha)_{\alpha\in I}$ of X there is a finite subcovering $(V_\beta)_{\beta\in J}$.

We may alternately define compactness by the statement, "X is compact if for each open covering $(U_\alpha)_{\alpha\in I}$ of X there is a finite subset of indices $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that the collection $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n}$ covers X."

Definition 2.5 A subset C of a topological space X is said to be *compact*, if C is a compact topological space in the relative topology.

A topological space C may be a subspace of two distinct larger topological spaces X and Y. In this event the relative topology of C is the same whether we regard C as a subspace of X or of Y, and, consequently, the assertion C is compact depends only on C and its topology. We may relate the compactness of a subspace C of a topological space X to the topology of X by means of the following theorem.

Theorem 2.6 A subset C of a topological space X is compact if and only if for each open covering $(U_\alpha)_{\alpha\in I}$ of C, U_α open in X, there is a finite subcovering $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n}$ of C.

Proof. Let C be compact and let $(U_\alpha)_{\alpha\in I}$ be an open covering of C. Thus $C \subset \bigcup_{\alpha\in I} U_\alpha$, hence $C = \bigcup_{\alpha\in I} (U_\alpha \cap C)$, so that the family $(U_\alpha \cap C)_{\alpha\in I}$ is a covering of the topological space C by relatively open sets $U_\alpha \cap C$. Since C is compact, there is a finite collection of indices $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that $U_{\alpha_1} \cap C$, $U_{\alpha_2} \cap C$, $\ldots, U_{\alpha_n} \cap C$ covers C; that is $C = \bigcup_{i=1}^{n} (U_{\alpha_i} \cap C)$. Consequently, $C \subset \bigcup_{i=1}^{n} U_{\alpha_i}$.

Conversely, suppose that for each open covering $(U_\alpha)_{\alpha\in I}$ of C, there is a finite subcovering $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n}$ of C. We must show that given any covering of the topological space C by a family $(V_\beta)_{\beta\in J}$ of relatively open subsets of C, there is a finite subcovering. For each $\beta \in J$, since V_β is relatively open in C, there is an open subset U_β of X such that $V_\beta = U_\beta \cap C$. But $C = \bigcup_{\beta\in J} V_\beta$, therefore $C \subset \bigcup_{\beta\in J} U_\beta$ and $(U_\beta)_{\beta\in J}$ is a covering of the subset C by open sets. By our hypothesis, there is a finite subcovering $U_{\beta_1}, U_{\beta_2}, \ldots, U_{\beta_m}$ of C. Thus

$$C \subset \bigcup_{i=1}^{m} U_{\beta_i} \quad \text{and} \quad C = \left(\bigcup_{i=1}^{m} U_{\beta_i}\right) \cap C = \bigcup_{i=1}^{m} (U_{\beta_i} \cap C) = \bigcup_{i=1}^{m} V_{\beta_i}.$$

Hence the covering $(V_\beta)_{\beta\in J}$ of C by relatively open sets has a finite subcovering $V_{\beta_1}, V_{\beta_2}, \ldots, V_{\beta_m}$.

Compactness may be characterized in terms of neighborhoods.

Theorem 2.7 A topological space X is compact if and only if, whenever for each $x \in X$ a neighborhood N_x of x is given, there is a finite number of points $x_1, x_2, \ldots, x_n$ of X such that $X = \bigcup_{i=1}^{n} N_{x_i}$.

Proof. Suppose X is compact. Let there be given for each $x \in X$ a neighborhood N_x of x. For each x, there is an open set U_x such that $x \in U_x \subset N_x$ and consequently the family $(U_x)_{x \in X}$ is an open covering of X. Since X is compact there is a finite subcovering $U_{x_1}, U_{x_2}, \ldots, U_{x_n}$. But $U_{x_i} \subset N_{x_i}$ for each i, whence $N_{x_1}, N_{x_2}, \ldots, N_{x_n}$ covers X.

Conversely, suppose that whenever, for each $x \in X$ a neighborhood N_x of x is given, there is a finite number of points $x_1, x_2, \ldots, x_n$ of X such that $X = \bigcup_{i=1}^{n} N_{x_i}$. Let $(U_\alpha)_{\alpha \in I}$ be an open covering of X. Then, for each $x \in X$, there is an $\alpha = \alpha(x)$ such that $x \in U_\alpha$, and therefore $N_x = U_\alpha$ is a neighborhood of x. By our hypothesis, there are points $x_1, x_2, \ldots, x_n$ of X such that $N_{x_i} = U_{\alpha(x_i)}$, $i = 1, 2, \ldots, n$, covers X, and hence X is compact.

In terms of closed sets, we have:

Theorem 2.8 A topological space is compact if and only if whenever a family $(F_\alpha)_{\alpha \in I}$ of closed sets is such that $\bigcap_{\alpha \in I} F_\alpha = \emptyset$ then there is a finite subset of indices $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that $\bigcap_{i=1}^{n} F_{\alpha_i} = \emptyset$.

Proof. Suppose X is compact and a family $(F_\alpha)_{\alpha \in I}$ of closed sets is given such that $\bigcap_{\alpha \in I} F_\alpha = \emptyset$. Then

$$\bigcup_{\alpha \in I} C(F_\alpha) = C\left(\bigcap_{\alpha \in I} F_\alpha\right) = X,$$

so that $(C(F_\alpha))_{\alpha \in I}$ is an open covering of X. Hence there is a finite subcovering $C(F_{\alpha_1}), C(F_{\alpha_2}), \ldots, C(F_{\alpha_n})$. Therefore

$$\bigcap_{i=1}^{n} F_{\alpha_i} = C\left(\bigcup_{i=1}^{n} C(F_{\alpha_i})\right) = \emptyset.$$

Conversely, suppose that for each family $(F_\alpha)_{\alpha \in I}$ of closed sets such that $\bigcap_{\alpha \in I} F_\alpha = \emptyset$ there is a finite subset of indices $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that $\bigcap_{i=1}^{n} F_{\alpha_i} = \emptyset$. Let $(O_\beta)_{\beta \in J}$ be an open covering of X. Then $(C(O_\beta))_{\beta \in J}$ is a family of closed sets such that $\bigcap_{\beta \in J} C(O_\beta) = \emptyset$. Thus $\bigcap_{i=1}^{n} C(O_{\alpha_i}) = \emptyset$ and $O_{\alpha_1}, O_{\alpha_2}, \ldots, O_{\alpha_n}$ is a finite subcovering.

The following theorem states that the continuous image of a compact set is compact and yields the corollary that compactness is a topological property.

Theorem 2.9 Let $f: X \to Y$ be continuous and let A be a compact subset of X. Then $f(A)$ is a compact subset of Y.

Proof. Let $(U_\alpha)_{\alpha \in I}$ be an open covering of $f(A)$. Thus $f(A) \subset \bigcup_{\alpha \in I} U_\alpha$ and consequently $A \subset \bigcup_{\alpha \in I} f^{-1}(U_\alpha)$ so that $(f^{-1}(U_\alpha))_{\alpha \in I}$ is a covering of A. Since f is continuous, $f^{-1}(U_\alpha)$ is an open subset of X for each $\alpha \in I$ and therefore $(f^{-1}(U_\alpha))_{\alpha \in I}$ is an open covering of A. A is compact, thus there is a finite subcovering $f^{-1}(U_{\alpha_1}), f^{-1}(U_{\alpha_2}), \ldots, f^{-1}(U_{\alpha_n})$ of A. But $A \subset f^{-1}(U_{\alpha_1}) \cup f^{-1}(U_{\alpha_2}) \cup \ldots \cup f^{-1}(U_{\alpha_n})$ implies that $f(A) \subset U_{\alpha_1} \cup U_{\alpha_2} \cup \ldots \cup U_{\alpha_n}$. $(U_\alpha)_{\alpha \in I}$ was an arbitrary open covering of $f(A)$, whence by Theorem 2.6, we have shown that $f(A)$ is compact.

Corollary 2.10 Let the topological spaces X and Y be homeomorphic. Then X is compact if and only if Y is compact.

Not every subset of a compact space is itself compact. We shall see that the closed interval $[0, 1]$ is compact, whereas of the open interval $(0, 1)$ is not compact. To show that $(0, 1)$ is not compact, it suffices to find one open covering of $(0, 1)$ that does not have a finite subcovering. To this end, for each integer $n = 3, 4, 5, \ldots$, let $U_n = \left(\frac{1}{n}, 1 - \frac{1}{n}\right)$. Then $(U_n)_{n=3,4,5,\ldots}$ is an open covering of $(0, 1)$. On the other hand,

for each integer $k > 3$ we have $\frac{1}{k} \notin \bigcup_{n=3}^{k} U_n$. Thus the union of every finite subcollection of $(U_n)_{n=3,4,5,\ldots}$ must fail to contain some point of $(0, 1)$, and hence there is no finite subcovering of $(U_n)_{n=3,4,5,\ldots}$.

We do, however, have this result.

Theorem 2.11 Let X be compact. Then each closed subset of X is compact.

Proof. Let F be a closed subset of the compact space X. If $(U_\alpha)_{\alpha\in I}$ is an open covering of F, then by adjoining the open set $O = C(F)$ to the family $(U_\alpha)_{\alpha\in I}$ we obtain an open covering $(V_\beta)_{\beta\in J}$ of X. Since X is compact there is a finite subcovering $V_{\beta_1}, V_{\beta_2}, \ldots,$ V_{β_m} of X. But each V_{β_i} is either equal to a U_α for some $\alpha \in I$ or equal to O. If O occurs among $V_{\beta_1}, V_{\beta_2}, \ldots, V_{\beta_m}$ we may delete it to obtain a finite collection of the U_α's that covers $F = C(O)$.

Thus, in a compact space, for each subset the property of being closed implies the property of being compact. In a Hausdorff space, the converse is also true.

Theorem 2.12 Let X be a Hausdorff space. If a subset F of X is compact, then F is closed.

Proof. We shall show that $O = C(F)$ is open by showing that for each point $z \in O$ there is a neighborhood U of z contained in O, or equivalently, for which $U \cap F = \emptyset$. To this end, with $z \in O$ fixed, by the Hausdorff property, we may choose for each point $x \in F$ an open neighborhood U_x of z and an open neighborhood V_x of x such that $U_x \cap V_x = \emptyset$. The family $(V_x)_{x\in F}$ is an open covering of F. Since F is compact, there is a finite subcovering $V_{x_1}, V_{x_2}, \ldots,$ V_{x_n} of F. The intersection $U = U_{x_1} \cap U_{x_2} \cap \cdots \cap U_{x_n}$ is an intersection of a finite set of neighborhoods of z and is therefore a neighborhood of z. Furthermore, U cannot intersect F since it does not intersect each element $V_{x_1}, V_{x_2}, \ldots, V_{x_n}$ of a covering of F. Thus $U \subset O$, from which it follows that O is a neighborhood of each of its points and $F = C(O)$ is closed.

Although the last two theorems are not precisely converses of each other, since the conditions imposed on the space X are different, we obtain:

Corollary 2.13 Let X be a compact Hausdorff space. Then a subset F of X is compact if and only if it is closed.

A significant result that follows from these last two theorems is:

Theorem 2.14 Let $f: X \to Y$ be a one-one continuous mapping of the compact space X onto a Hausdorff space Y. Then f is a homeomorphism.

Proof. We define $g: Y \to X$ by setting $g(y) = x$ if $f(x) = y$, so that f and g are inverse functions. It remains to prove that g is continuous. We shall prove this by proving that for each closed subset F of X, $g^{-1}(F)$ is a closed subset of Y. Given a closed subset F of X, by Theorem 2.11, F is compact. Hence $f(F) = g^{-1}(F)$ is a compact subset of Y. By Theorem 2.12, $g^{-1}(F)$ is a closed subset of Y. Thus, g is continuous and f is a homeomorphism.

Theorems 2.11 and 2.12 will be extremely useful in proving the main result of the next section, namely, that a subset of the real line is compact if and only if it is closed and "bounded." Since the real line is a Hausdorff space, Theorem 2.12 immediately implies that a compact subset of the real line is closed. Furthermore, the family of open intervals $(U_n)_{n=1,2,3,\ldots}$, where $U_n = (-n, n)$, is an open covering of each subset F of the real line. In the event that F is compact, it is covered by a finite subcovering of $(U_n)_{n=1,2,3,\ldots}$ and hence "bounded"; that is, contained in $(-K, K)$ for a sufficiently large K. Thus,, the "only if" part of this theorem is easily derived. With regard to the "if" part, that is, a closed and bounded subset of the real line is compact, we shall use the results of this section to reduce the proof to the particular proposition that the unit

interval [0, 1] is compact. The basis for this reduction is the reasoning: a closed and bounded subset F of the real line is a closed subset of a sufficiently large closed interval $[-K, K]$. Thus, to prove F is compact, it suffices, by Theorem 2.11, to prove that $[-K, K]$ is compact. But $[-K, K]$ is homeomorphic to $[0, 1]$; thus it suffices to prove that $[0, 1]$ is compact.

Exercises

1. Let $B = (0, 1]$. Set $A_1 = (\frac{1}{2}, 2)$ and
$$A_n = \left(\frac{1}{n+1}, \frac{1}{n-1}\right)$$
for each integer $n > 1$. Prove that $(A_n)_{n \in N}$ is an open covering of B, where N is the set of natural numbers, and that no finite subset of $(A_n)_{n \in N}$ covers B.
2. Let $F:[a, b] \to [c, d]$ be a continuous increasing function such that $f(a) = c, f(b) = d$. Prove that f is a homeomorphism.
3. A subset F of R^n is said to be *bounded* if there is some positive number K such that for each $x = (x_1, x_2, \ldots, x_n) \in F$, $|x_i| \leq K$ for $i = 1, 2, \ldots, n$. Assume that the unit cube I^n is compact and using this assumption prove that a subset F of R^n is compact if and only if it is closed and bounded.
4. Let X be compact and $f:X \to R$ continuous. Prove that $f(X)$ is closed and bounded. Let M be the least upper bound of $f(X)$ and m the greatest lower bound of $f(X)$. Prove that there are points x_1, x_2 in X such that $f(x_1) = M$ and $f(x_2) = m$.
5. Prove that every finite subset of a topological space is compact.
6. Let $(U_\alpha)_{\alpha \in I}$ be an open covering of $[0, 1]$. Define a subset P of $[0, 1]$ as follows: a point x is in P if and only if there is a finite collection $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_m}$ of elements of $(U_\alpha)_{\alpha \in I}$ that covers $[0, x]$. Prove that if $x \in P$, then there is a neighborhood O of x such that $O \subset P$ and that therefore P is open. Prove that if $x \notin P$, then there is a neighborhood O of x such that $O \cap P = \emptyset$ and therefore P is closed. Conclude that $P = [0, 1]$ and that therefore $[0, 1]$ is compact.

7. Let X be a topological space. A family $(F_\alpha)_{\alpha \in I}$ of subsets of X is said to have the *finite intersection property* if for each finite subset J of I, $\bigcap_{\alpha \in J} F_\alpha \neq \emptyset$. Prove that X is compact if and only if for each family $(F_\alpha)_{\alpha \in I}$ of closed subsets of X that has the finite intersection property, we have $\bigcap_{\alpha \in I} F_\alpha \neq \emptyset$.

3 Compact Subsets of the Real Line

Definition 3.1 A subset A of the real line is said to be *bounded* if there is a number $K > 0$ such that for each $x \in A$, $|x| \leq K$.

Thus A is bounded if and only if A is contained in some closed interval $[-K, K]$, $K > 0$. In particular, each closed interval $[a, b]$ is bounded, for if $K = \text{maximum } \{|a|, |b|\}$ then $[a, b] \subset [-K, K]$.

Theorem 3.2 A subset A of the real line is compact if and only if A is closed and bounded.

Proof. First, suppose A is compact. The real line satisfies the Hausdorff axiom, so, by Theorem 2.12, A is closed. For each positive integer n, let $O_n = (-n, n)$. $R = \bigcup_{n \in N} O_n$, where N is the set of natural numbers. Therefore $(O_n)_{n \in N}$ is an open covering of A. Since A is compact, $A \subset O_{n_1} \cup O_{n_2} \cup \cdots \cup O_{n_q}$. If we set $k = \text{maximum } \{n_1, n_2, \ldots, n_q\}$ then $O_{n_i} \subset O_k$ for $i = 1, 2, \ldots, q$, whence $A \subset O_k = (-k, k)$. Thus $A \subset [-k, k]$ and A is bounded.

To prove the converse, we first establish the following special case:

Lemma 3.3 The closed interval $[0, 1]$ is compact.

Proof. Let $(O_\alpha)_{\alpha \in I}$ be a covering of $[0, 1]$ by open sets. Assume that there is no finite subcovering of $(O_\alpha)_{\alpha \in I}$. In this event, at least one of the two closed intervals $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$ cannot be covered by

a finite subcollection of the family $(O_\alpha)_{\alpha\in I}$. Let $[a_1, b_1]$ denote one of these two intervals of length $\frac{1}{2}$ that cannot be covered by a finite subcollection of the family $(O_\alpha)_{\alpha\in I}$. We may now divide the interval $[a_1, b_1]$ into the two subintervals

$$\left[a_1, \frac{a_1 + b_1}{2}\right]$$

and

$$\left[\frac{a_1 + b_1}{2}, b_1\right]$$

of length $\frac{1}{4}$ and assert that at least one of these two intervals cannot be covered by a finite subcollection of the family $(O_\alpha)_{\alpha\in I}$. Let $[a_2, b_2]$ denote one of these two intervals of length $\frac{1}{4}$ that has the property that it cannot be covered by a finite subcollection of the family $(O_\alpha)_{\alpha\in I}$. We shall now proceed to define a sequence $[a_0, b_0]$, $[a_1, b_1]$, $[a_2, b_2]$, . . . , $[a_n, b_n]$, . . . of such intervals. Assume that for $r = 0, 1, 2, \ldots, n$ we have defined intervals $[a_r, b_r]$ such that:

1. $[a_0, b_0] = [0, 1]$;
2. $b_r - a_r = \dfrac{1}{2^r}$ for $r = 0, 1, \ldots, n$;
3. for $r = 0, 1, \ldots, n-1$, either $[a_{r+1}, b_{r+1}] = \left[a_r, \dfrac{a_r + b_r}{2}\right]$ or $[a_{r+1}, b_{r+1}] = \left[\dfrac{a_r + b_r}{2}, b_r\right]$;
4. for each $r = 0, 1, \ldots, n$, no finite subcollection of $(O_\alpha)_{\alpha\in I}$ covers $[a_r, b_r]$.

We then define $[a_{n+1}, b_{n+1}]$ in the following manner. In view of (4) at least one of the two intervals

$$\left[a_n, \frac{a_n + b_n}{2}\right], \left[\frac{a_n + b_n}{2}, b_n\right]$$

cannot be covered by a finite subcollection of $(O_\alpha)_{\alpha\in I}$. Denote by $[a_{n+1}, b_{n+1}]$ whichever of these two intervals cannot be covered by a finite subcollection of $(O_\alpha)_{\alpha\in I}$, (if neither can be, we may agree that $[a_{n+1}, b_{n+1}]$ is the first of the two). Then conditions (2), (3), and (4) will also hold for $[a_{n+1}, b_{n+1}]$. It follows by induction that we may

define a sequence $[a_0, b_0]$, $[a_1, b_1]$, $[a_2, b_2]$, . . . of such intervals for which conditions (1) through (4) are true.

By conditions (3), $a_n \leqq a_{n+1} \leqq b_{n+1} \leqq b_n$. It follows that for each pair of positive integers m and n, $a_m \leqq b_n$. Thus each b_n is an upper bound of the set $\{a_0, a_1, a_2, \ldots\}$. Let a be the least upper bound of this set. Then $a \leqq b_n$ for each n, and hence a is a lower bound of the set $\{b_0, b_1, b_2, \ldots\}$. Let b be the greatest lower bound of the latter set. We therefore have $a \leqq b$. But, by the definition of a and b, we must have $a_n \leqq a \leqq b \leqq b_n$ for each n, whence by condition (2), $b - a \leqq \dfrac{1}{2^n}$ for each n and we conclude that $a = b$. We are now in a position to obtain a contradiction to condition (4), from which it will follow that our assumption that there is no finite subcovering of $[0, 1]$ is untenable.

$(O_\alpha)_{\alpha\in I}$ covers $[0, 1]$ and $a = b \in [0, 1]$. Therefore $a \in O_\beta$ for some $\beta \in I$. Since O_β is open there is an $\varepsilon > 0$ such that $S(a; \varepsilon) \subset O_\beta$. Let us choose the positive integer N large enough so that $\dfrac{1}{2^N} < \varepsilon$. Then $b_N - a_N < \varepsilon$. Now $a = b \in [a_N, b_N]$. Therefore, $a - a_N < \dfrac{1}{2^N} < \varepsilon$ and $b - b_N < \dfrac{1}{2_N} < \varepsilon$. Consequently, $[a_N, b_N] \subset S(a; \varepsilon) \subset O_\beta$. Thus $[a_N, b_N]$ may be covered by a finite subcollection (namely, one!) of the family $(O_\alpha)_{\alpha\in I}$. Therefore the assumption that no finite subcollection of $(O_\alpha)_{\alpha\in I}$ covers $[0, 1]$ leads to a contradiction and we must conclude that $[0, 1]$ is compact.

It can be seen that the gist of the above argument is that if no finite subcollection of $(O_\alpha)_{\alpha\in I}$ covers $[0, 1]$, then no finite subcollection of $(O_\alpha)_{\alpha\in I}$ covers a sequence of subintervals whose lengths approach zero, whereas on the other hand if the length of one of these subintervals is small enough it is contained in some O_β.

Since each closed interval $[a, b]$ is homeomorphic to the closed interval $[0, 1]$ and compactness is a topological property, we obtain:

Corollary 3.4 Each closed interval $[a, b]$ is compact.

We now return to the part of Theorem 3.2 in which we prove that, if A is a closed and bounded subset of the real line, then A is compact.

Since A is bounded and closed, A is a closed subset of a closed interval $[-K, K]$ for some $K > 0$. But $[-K, K]$ is a compact space and therefore, by Theorem 2.11, A is compact. The proof of Theorem 3.2 is thus completed.

Exercises

1. Using the method of subdivision of Lemma 3.3, prove that the unit square I^2 is a compact subset of the plane and in general that the unit n-cube I^n is a compact subspace of R^n.
2. Let X be a compact space and $(F_n)_{n=1,2,3,\ldots}$ a sequence of nonempty closed subsets of X such that $F_{n+1} \subset F_n$ for each n. Prove that $\bigcap_{n=1} F_n \neq \emptyset$.
3. Let $f: [a, b] \to R$ be continuous. Prove that the set $f([a, b])$ has both a least upper bound M and a greatest lower bound m and that there are points $u, v \in [a, b]$ such that $f(u) = M, f(v) = m$.

4 Products of Compact Spaces

The fundamental result of this section is that the product of two compact spaces is itself compact. We shall establish this fact with the aid of the next lemma, which relates compactness to coverings by members of a base for the open sets. Let us recall that a base for the open sets of a topological space Z is a collection $\mathcal{B}$ of open subsets with the property that each open subset of Z is a union of members of the collection $\mathcal{B}$.

Lemma 4.1 Let $\mathcal{B}$ be a base for the open sets of a topological space Z. If, for each covering $(B_\beta)_{\beta \in J}$ of Z by members of $\mathcal{B}$, there is a finite subcovering, then Z is compact.

Proof. We must show that, if each covering of Z by "basic" open sets has a finite subcovering, then each open covering $(O_\alpha)_{\alpha \in I}$ of Z has a finite subcovering. For each $\alpha \in I$, O_α is a union of members of $\mathcal{B}$. Let J be an indexing set for all the basic sets B_β that occur in the expression of some O_α as a union of members of $\mathcal{B}$. Thus $\bigcup_{\alpha \in I} O_\alpha \subset \bigcup_{\beta \in J} B_\beta$ and hence $(B_\beta)_{\beta \in J}$ is a covering of Z by members of $\mathcal{B}$. It follows from our hypothesis that there is a finite subcovering $B_{\gamma_1}, B_{\gamma_2}, \ldots, B_{\gamma_n}$ of Z. Since each B_{γ_i} occurs in the expression of some O_{α_i}, $\alpha_i \in I$, as a union of members of $\mathcal{B}$, $B_{\gamma_i} \subset O_{\alpha_i}$. Consequently, $O_{\alpha_1}, O_{\alpha_2}, \ldots, O_{\alpha_n}$ must cover Z and therefore Z is compact.

Lemma 7.1 of Chapter III states that if X and Y are topological spaces, then a base for open sets of $X \times Y$ is the collection of sets of the form $U \times V$, where U is open in X and V is open in Y.

Theorem 4.2 Let X and Y be compact topological spaces; then $X \times Y$ is compact.

Proof. By virtue of Lemma 4.1 it suffices to prove that each covering of $X \times Y$ by sets of the form $U \times V$, U open in X, V open in Y, has a finite subcovering. Let $(U_\alpha \times V_\alpha)_{\alpha \in I}$ be such a covering. As an aid to understanding the proof, let us view the product $X \times Y$ as pictured in Figure 32, where a point $(x, y) \in X \times Y$ lies over the point $x \in X$ and level with the point $y \in Y$. In particular, for each $x_0 \in X$, the subset Y_{x_0} of $X \times Y$ consisting of all points (x_0, y), $y \in Y$, may be thought of as the collection of points lying over x_0. The open covering $(U_\alpha \times V_\alpha)_{\alpha \in I}$ is necessarily an open covering of Y_{x_0}. But Y_{x_0} is homeomorphic to Y and hence compact. We may therefore find a finite subset I_{x_0} of I such that $(U_\alpha \times V_\alpha)_{\alpha \in I_{x_0}}$ covers Y_{x_0} [this finite covering of Y_{x_0} is portrayed by the small rectangles in Figure 32]. We may also assume that $x_0 \in U_\beta$ for each $\beta \in I_{x_0}$, for otherwise we may delete $U_\beta \times V_\beta$ and still cover Y_{x_0}. The set $U^*_{x_0} = \bigcap_{\alpha \in I_{x_0}} U_\alpha$ is a finite intersection of open sets con-

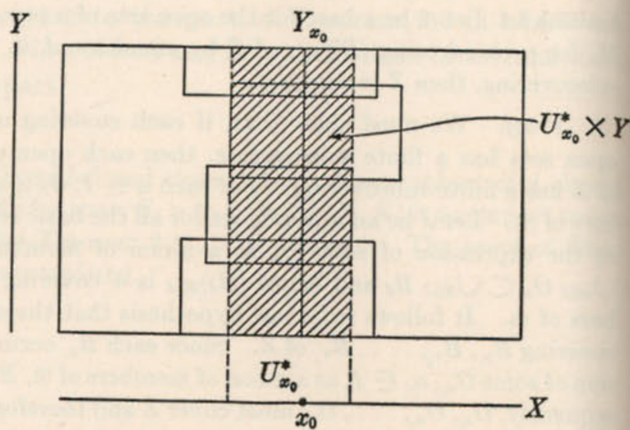


Figure 32

taining x_0 and is therefore an open set containing x_0. We assert that $(U_\alpha \times V_\alpha)_{\alpha \in I_{x_0}}$ is an open covering of $U^*_{x_0} \times Y$. For, suppose $(x, y) \in U^*_{x_0} \times Y$. The point (x_0, y) is in $U_\beta \times V_\beta$ for some $\beta \in I_{x_0}$. Since $x \in U^*_{x_0}$, $x \in U_\alpha$ for all $\alpha \in I_{x_0}$. It follows that $(x, y) \in U_\beta \times V_\beta$, $\beta \in I_{x_0}$, proving our assertion.

Now $(U^*_x)_{x \in X}$ is an open covering of the compact space X, hence there is a finite subcovering $U^*_{x_1}, U^*_{x_2}, \ldots, U^*_{x_n}$ of X. Let us set $I^* = I_{x_1} \cup I_{x_2} \cup \ldots \cup I_{x_n}$ and show that the finite family $(U_\alpha \times V_\alpha)_{\alpha \in I^*}$ is a covering of $X \times Y$. Given a point $(x, y) \in X \times Y$, $x \in U^*_{x_i}$ for some x_i so that $(x, y) \in U^*_{x_i} \times Y$. By our previous assertion $(x, y) \in U_\beta \times V_\beta$ for some $\beta \in I_{x_i}$ which certainly implies that $(x, y) \in U_\alpha \times V_\alpha$ for some $\alpha \in I^*$. We have thus established that $(U_\alpha \times V_\alpha)_{\alpha \in I^*}$ is a finite subcovering and that therefore $X \times Y$ is compact.

If $X_1, X_2, \ldots, X_n$ are topological spaces, one may distinguish between $\prod_{i=1}^{n} X_i$ and $\left(\prod_{i=1}^{n-1} X_i\right) \times X_n$, for the points of the first space are n-tuples $(x_1, x_2, \ldots, x_n)$, whereas the points of the second space are ordered pairs $((x_1, x_2, \ldots, x_{n-1}), x_n)$ whose first elements are $(n-1)$-tuples. Nevertheless, these two spaces are certainly homeomorphic [the obvious homeomor-

phism carries a point $(x_1, x_2, \ldots, x_n)$ into $((x_1, x_2, \ldots, x_{n-1}), x_n)$], hence by induction on n we obtain:

Corollary 4.3 Let $X_1, X_2, \ldots, X_n$ be compact topological spaces. Then $\prod_{i=1}^{n} X_i$ is also compact.

Let us recall that the unit n-cube I^n is the subset of R^n consisting of all points $x = (x_1, x_2, \ldots, x_n)$ such that $0 \leqq x_i \leqq 1$ for $i = 1, 2, \ldots, n$. As a subspace of R^n, I^n has the same topology as the product space $I \times I \times \ldots \times I$ (n-factors). Since $I = [0, 1]$ is compact, as a special case of Corollary 4.3 we have:

Corollary 4.4 The unit n-cube I^n is compact.

As a generalization of boundedness on the real line we make the following definition.

Definition 4.5 A subset A of R^n is said to be *bounded* if there is a real number $k > 0$ such that for each point $x = (x_1, x_2, \ldots, x_n) \in A$, $|x_i| \leqq k$ for $i = 1, 2, \ldots, n$.

Let $0_n = (0, 0, \ldots, 0)$ be the origin of R^n. To say that a subset A of R^n is bounded is equivalent to the assertion that $A \subset S(0_n, K)$ for some $K > 0$.

Theorem 4.6 A subset A of R^n is compact if and only if A is closed and bounded.

Proof. The proof that if A is compact then A is closed and bounded is similar to the proof of this fact for a subset of the real line as presented in the preceding section. Conversely, we shall first show that each closed "cube" is compact. The collection of points $x = (x_1, x_2, \ldots, x_n)$ in R^n such that $|x_i| \leqq K$ for $i = 1, 2, \ldots, n$, which we shall denote by M_K, is a "cube" of width $2K$ with center

at the origin. M_K is homeomorphic to the unit n-cube I^n, for the function $F: I^n \to M_K$ defined by

$$F(x_1, x_2, \ldots, x_n) = (2Kx_1 - K, 2Kx_2 - K, \ldots, 2Kx_n - K)$$

is easily seen to be a homeomorphism (Theorem 2.14). Since I^n is compact, M_K is compact. Now suppose A is closed and bounded; then A is a closed subset of the compact cube M_K for some K, whence A is compact.

Exercises

1. Let S be the set $[0, 1]$ and make the definition: a subset F of S is closed if it is finite. Prove that this definition of closed set yields a topology for S. Show that S with this topology is connected, arcwise connected, and compact, but that S is not a Hausdorff space. Show that each subset of S is compact and that therefore there are compact subsets of S that are not closed.
2. A topological space X is said to be *locally compact* if each point $x \in X$ has at least one compact neighborhood. Prove that the real line and R^n are locally compact.
3. Let X be a topological space and x^* a point of X. Assume a base for the system of neighborhoods of x^* consists of the complements of compact subsets of $X - \{x^*\}$. Prove X is compact. Prove that if in addition $X - \{x^*\}$ is a locally compact Hausdorff space, then X is a compact Hausdorff space. Given a locally compact Hausdorff space Y, show that Y is a subspace of a compact Hausdorff space that contains one more point than Y does called the *one-point compactification of* Y.

5 Compact Metric Spaces

A metric space (X, d) is said to be *compact* or is called a *compactum* if its associated topological space is compact. In this section we shall derive certain properties of compact metric

spaces. A basis result is that a metric space is compact if and only if every infinite subset has at least one "point of accumulation."

Definition 5.1 Let X be a topological space and A a subset of X. A point $a \in X$ is called an *accumulation point* of A if each neighborhood of a contains infinitely many distinct points of A.

In referring to the accumulation points of a set A, care must be taken to specify of which topological space A is to be considered a subset. For example, in the real line R, the subset $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots\right\}$ has the accumulation point 0, whereas in the topological space $(0, +\infty)$, the same set A has no accumulation points.

A seemingly weaker characterization of accumulation point may be used in a Hausdorff space and, consequently, also in a metric space.

Lemma 5.2 Let X be a Hausdorff space and A a subset of X. A point $a \in X$ is an accumulation point of A if and only if each neighborhood of a contains a point of A distinct from a.

Proof. The "only if" part of the theorem is immediate. Suppose, conversely, that a is not an accumulation point of A. Then there is a neighborhood N of a that contains at most a finite collection $\{a_1, a_2, \ldots, a_p\}$ of points of A distinct from a. For each of these points a_i, $i = 1, 2, \ldots, p$, we can find neighborhoods U_i of a and neighborhoods V_i of a_i such that $U_i \cap V_i = \emptyset$. Then $N \cap U_1 \cap U_2 \cap \cdots \cap U_p$ is a neighborhood of a that contains no points of A other than possibly a. The proof of the "if" part of Lemma 5.2 is now complete.

As suggested in the above argument, whether or not a point a is an accumulation point of a subset A does not depend on whether a is a point of A. In fact, it is a direct consequence of the definition that a point a is an accumulation

point of a subset A if and only if a is an accumulation point of $A - \{a\}$. If N is a neighborhood of a point a, the subset $N - \{a\}$ is called a *deleted neighborhood* of a. Lemma 5.2 thus states that in a Hausdorff space a point a is an accumulation point of a subset A if and only if each deleted neighborhood of a contains a point of A. A point $b \in A$ that is not an accumulation point of A must therefore possess a neighborhood W such that $W \cap A = \{b\}$. Such a point is called an *isolated* point of A.

For metric spaces the distance function may be used to characterize accumulation points.

Lemma 5.3 Let A be a subset of a metric space (X, d). A point $a \in X$ is an accumulation point of A if and only if $d(a; A - \{a\}) = 0$.

Proof. If $d(a; A - \{a\}) > 0$ then choosing ε so that $0 < \varepsilon < d(a; A - \{a\})$, $S(a; \varepsilon) \cap A$ is either empty or $\{a\}$ and hence a is not an accumulation point of A. Conversely, if a is not an accumulation point of A then there is a neighborhood N, and hence a spherical neighborhood $S(a; \varepsilon)$, which has no points in common with $A - \{a\}$. Thus $d(a; A - \{a\}) \geqq \varepsilon > 0$.

Theorem 5.4 Let X be a compact Hausdorff space; then each infinite subset A of X has at least one point of accumulation in X.

Proof. We shall prove that, if a subset K of X has no accumulation points, then K is finite. Since K has no accumulation points, for each point $x \in X$, we can find a neighborhood N_x of x such that either $N_x \cap K = \emptyset$ or $N_x \cap K = \{x\}$. X is compact, hence there are points $x_1, x_2, \ldots, x_m$ such that $N_{x_1}, N_{x_2}, \ldots, N_{x_m}$ cover X. It follows that $K \subset \{x_1, x_2, \ldots, x_m\}$ and K is finite.

In the event that Theorem 5.4 is used where X is a compact subspace of a larger topological space, it is important to realize that the theorem states that the accumulation point is in X. In particular, if X is a closed and bounded subset of R^n, so that X is compact, each infinite subset A of X has an accumulation point in X.

We have now proved half of the theorem that states that a metric space is compact if and only if each infinite subset has a point of accumulation. The next lemma will be needed to prove the other half of this theorem.

Lemma 5.5 Let (X, d) be a metric space such that every infinite subset of X has at least one accumulation point in X. Then, for each positive integer n, there is a finite set of points $x_1^n, x_2^n, \ldots, x_p^n$ of X such that the collection of open spheres

$$S\left(x_1^n; \frac{1}{n}\right), S\left(x_2^n; \frac{1}{n}\right), \ldots, S\left(x_p^n; \frac{1}{n}\right)$$

covers X.

Proof. Suppose there were an integer n such that no finite collection of spheres of radius $\frac{1}{n}$ covered X. Choose a point $x_1 \in X$. $S\left(x_1; \frac{1}{n}\right)$ certainly does not cover X, hence there is a point $x_2 \in X$ such that $x_2 \notin S\left(x_1; \frac{1}{n}\right)$. $S\left(x_1; \frac{1}{n}\right) \cup S\left(x_2; \frac{1}{n}\right)$ does not cover X, hence there is a point $x_3 \in X$ such that $x_3 \notin S\left(x_1; \frac{1}{n}\right) \cup S\left(x_2; \frac{1}{n}\right)$. Continuing in this way we may construct a sequence $x_1, x_2, \ldots, x_k, \ldots$ of points of X such that for $k > 1$,

$$x_k \notin \bigcup_{i=1}^{k-1} S\left(x_i; \frac{1}{n}\right)$$

Thus

$$d(x_k, x_{k'}) \geqq \frac{1}{n}$$

if $k \neq k'$. It follows that the set $\{x_1, x_2, \ldots, x_k, \ldots\}$ is infinite and therefore has a point of accumulation $x \in X$. Therefore the neighborhood $S\left(x; \frac{1}{2n}\right)$ contains infinitely many points of $\{x_1, x_2, \ldots, x_k, \ldots\}$ and in particular contains two points $x_k, x_{k'}$ with $k \neq k'$. Since $x_k, x_{k'} \in S\left(x; \frac{1}{2n}\right)$, we obtain the contradiction $d(x_k, x_{k'}) < \frac{1}{n}$.

A similar argument yields the following result.

Lemma 5.6 Let (X, d) be a metric space such that each infinite subset of X has at least one point of accumulation. Then for each open covering $(O_\alpha)_{\alpha \in I}$ of X there is a positive number ε such that each open sphere $S(x; \varepsilon)$ is contained in an element O_β of this covering.

Proof. We shall suppose the result is false and obtain a contradiction. If the lemma is false, for each $n = 1, 2, \ldots$, there is an open sphere $S\left(x_n; \frac{1}{n}\right)$ such that $S\left(x_n; \frac{1}{n}\right) \not\subset O_\alpha$ for each $\alpha \in I$. Let $A = \{x_1, x_2, \ldots\}$. If A is finite, some point $x \in X$ occurs infinitely often in the sequence $x_1, x_2, \ldots$. Since $(O_\alpha)_{\alpha \in I}$ covers X, $x \in O_\beta$ for some $\beta \in I$. Also, O_β is open, hence there is a $\delta > 0$ such that $S(x; \delta) \subset O_\beta$. We may, however, choose n so that $\frac{1}{n} < \delta$ and $x_n = x$, in which case $S\left(x_n; \frac{1}{n}\right) = S\left(x; \frac{1}{n}\right) \subset O_\beta$, a contradiction. There remains the possibility that $A = \{x_1, x_2, \ldots\}$ is infinite. Thus A has at least one point of accumulation x. Again $x \in O_\beta$ for some $\beta \in I$ so that $S(x; \delta) \subset O_\beta$ for some $\delta > 0$. There are infinitely many points of A in the neighborhood $S\left(x; \frac{\delta}{2}\right)$ of x. Hence we may choose an n such that $\frac{1}{n} < \frac{\delta}{2}$ and $x_n \in S\left(x; \frac{\delta}{2}\right)$. We then have $S\left(x_n; \frac{1}{n}\right) \subset S(x; \delta) \subset O_\beta$, which is again a contradiction.

The number ε of Lemma 5.6 depends on the particular open covering $(O_\alpha)_{\alpha \in I}$ considered. Given the open covering $(O_\alpha)_{\alpha \in I}$, if the number ε has the property that for each $x \in X$, $S(x; \varepsilon) \subset O_\beta$ for some $\beta \in I$, then each number ε' with $0 < \varepsilon' < \varepsilon$ also has this property. The least upper bound of the set of numbers having this property is called the *Lebesque number*, ε_L, of the open covering $(O_\alpha)_{\alpha \in I}$. We may now state:

Corollary 5.7 Let (X, d) be a metric space such that each infinite subset of X has an accumulation point. Then each open covering $(O_\alpha)_{\alpha \in I}$ of X has a Lebesque number ε_L.

A topological space X is said to have the *Bolzano-Weierstrass property* if each infinite subset of X has at least one point of accumulation. We may now prove that every metric space that has the Bolzano-Weierstrass property is a compact metric space.

Theorem 5.8 Let (X, d) be a metric space that has the property that every infinite subset of X has at least one accumulation point. Then X is compact.

Proof. Let $(O_\alpha)_{\alpha \in I}$ be an open covering and let ε_L be its Lebesque number. Let us choose n so that $\frac{1}{n} < \varepsilon_L$. By Lemma 5.5 there is a finite set $\{x_1, x_2, \ldots, x_p\}$ of points of X such that the open spheres $S\left(x_1; \frac{1}{n}\right), S\left(x_2; \frac{1}{n}\right), \ldots, S\left(x_p; \frac{1}{n}\right)$ cover X. Furthermore, by Lemma 5.6, for each $i = 1, 2, \ldots, p$, there is a $\beta_i \in I$ such that $S\left(x_i; \frac{1}{n}\right) \subset O_{\beta_i}$. It follows that the collection $O_{\beta_1}, O_{\beta_2}, \ldots, O_{\beta_p}$ is a finite subcovering of $(O_\alpha)_{\alpha \in I}$.

We have now proved the main result of this section.

Theorem 5.9 Let (X, d) be a metric space. Each infinite subset of X has at least one accumulation point if and only if X is compact.

Having proved that a subspace X of Euclidean n-space R^n is compact if and only if it is closed and bounded, we may state:

Corollary 5.10 Let X be a subspace of R^n. Then the following three properties are equivalent.

1. X is compact.
2. X is closed and bounded.
3. Each infinite subset of X has at least one point of accumulation in X.

The existence, for each open covering of a compact metric space, of a Lebesque number has as a consequence the fact that each continuous function defined on a compact metric space is "uniformly" continuous.

Definition 5.11 Let $f:(X, d) \to (Y, d')$ be a function from a metric space (X, d) to a metric space (Y, d'). f is said to be *uniformly continuous* if, for each positive number ε, there is a $\delta > 0$, such that whenever

$$d(x, y) < \delta,$$

then

$$d'(f(x), f(y)) < \varepsilon.$$

If the function $g:X \to Y$ is continuous, then for each $x \in X$ and each $\varepsilon > 0$, there is a $\delta > 0$, where δ may depend on both the choice of x and ε, such that $d(x, a) < \delta$ implies $d'(g(x), g(a)) < \varepsilon$. If, however, g is uniformly continuous, then given $\varepsilon > 0$, the number δ may be used, at each point $x \in X$, that is, uniformly throughout X, to yield $d'(g(x), g(a)) < \varepsilon$ if $d(x, a) < \delta$. Thus:

Corollary 5.12 If $f:X \to Y$ is uniformly continuous, then f is continuous.

On the other hand a continuous function need not be uniformly continuous. As an example, consider $f:(0, 1] \to R$ defined by $f(x) = \frac{1}{x}$. Given $\varepsilon = 1$, we shall show that there does not exist a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < 1$. For given any $\delta > 0$ we can choose n large enough so that if $x = \frac{1}{n}$, $y = \frac{1}{n+1}$ we have

$$x - y = \frac{1}{n(n+1)} < \delta$$

whereas

$$\left|f\left(\frac{1}{n}\right) - f\left(\frac{1}{n+1}\right)\right| = 1.$$

In view of the next result, it should be noted that in this example the interval $(0, 1]$ is not compact.

Theorem 5.13 Let $f:(X, d) \to (Y, d')$ be a continuous function from a compact metric space X to a metric space Y. Then f is uniformly continuous.

Proof. Given $\varepsilon > 0$, for each $x \in X$, there is a $\delta_x > 0$ such that if $y \in S(x; \delta_x)$ then $f(y) \in S\left(f(x); \frac{\varepsilon}{2}\right)$. The collection $(S(x; \delta_x))_{x \in X}$ is an open covering of X. Since X is compact, this open covering has a Lebesque number. Let us choose δ to be a positive number less than this Lebesque number. If $z, z' \in X$ and $d(z, z') < \delta$ so that z and z' are in a sphere of radius less than δ, we have $z, z' \in S(x; \delta_x)$ for some $x \in X$. Consequently, $f(z), f(z') \in S\left(f(x); \frac{\varepsilon}{2}\right)$, whence $d'(f(z), f(z')) \leqq d'(f(z), f(x)) + d'(f(x), f(z')) < \varepsilon$.

Exercises

1. Let A be a subset of a Hausdorff space. Let A' denote the set of accumulation points of A and A^i denote the isolated points of A. Prove that $\overline{A} = A' \cup A^i$, $A' \cap A^i = \emptyset$.
2. In a metric space (X, d), a sequence $a_1, a_2, \ldots$ of points of X is called a *Cauchy sequence* if for each $\varepsilon > 0$ there is a positive integer N such that $d(a_n, a_m) < \varepsilon$ whenever $n, m > N$. A metric space X is called *complete* if every Cauchy sequence in X converges to a point of X. Prove that a compact metric space is complete.
3. In Euclidean n-space R^n, prove that every Cauchy sequence lies in a bounded closed subset of R^n. Use this fact to prove that R^n is complete.
4. Let (X, d) be a compact metric space. Prove that X is "bounded with respect to d"; that is, there is a positive number K such that $d(x, y) \leqq K$ for all $x, y \in X$.
5. Let (X, d) be a compact metric space and let $(F_\alpha)_{\alpha \in I}$ be a family of closed subsets of X such that $\bigcap_{\alpha \in I} F_\alpha = \emptyset$. Prove that there

is a positive number c such that for each $x \in X$, $d(x, F_\alpha) \geqq c$ for some $\alpha \in I$.

6. A subset A of a topological space X is called *dense* if $\overline{A} = X$. Let X be a compact metric space. Prove that there is a sequence $a_1, a_2, \ldots$ of points of X such that the set $A = \{a_1, a_2, \ldots\}$ is dense in X.

7. Let X be the set of continuous functions $f:[a, b] \to R$. Let $I:X \to R$ be defined by $I(f) = \int_a^b f(t)\, dt$. Define a distance function d on X by setting $d(f, g) = \underset{a \leqq t \leqq b}{\text{l.u.b.}} |f(t) - g(t)|$. Prove that I is uniformly continuous. Let $f_1, f_2, \ldots$ be a Cauchy sequence in (X, d). Prove that for each $t \in [a, b]$, $f_1(t), f_2(t), \ldots$ is a Cauchy sequence of real numbers. For each $t \in [a, b]$, denote by $f(t)$ the limit of this sequence. Prove that the function $f:[a, b] \to R$ so defined is continuous, that $\lim_n f_n = f$ in X, and therefore (X, d) is complete. Prove that (X, d) is not compact.

6 Compactness and the Bolzano-Weierstrass Property

Theorem 5.9, which states that a metric space is compact if and only if each infinite subset has at least one accumulation point, raises the question as to whether or not these two properties are equivalent in an arbitrary topological space. We already know that in general the first property implies the second. Since there are examples of topological spaces that are not compact, but in which each infinite subset has a point of accumulation, compactness is the stronger of the two properties. We might therefore think of the second property, which we have called the Bolzano-Weierstrass property, as a weaker type of compactness. To illustrate how much weaker the Bolzano-Weierstrass property is, we need to introduce the concept of countability.

Definition 6.1 A non-empty set X is said to be *countable* if there is an onto function $f:N \to X$, where N is the set of positive integers.

A finite set $X = \{x_1, x_2, \ldots, x_n\}$ is countable, for we may construct an onto function $f:N \to X$ by setting $f(i) = x_i$, $1 \leqq i \leqq n$, and defining $f(i)$ for $i > n$ arbitrarily, say $f(i) = x_1$, $i > n$. A countable set that is not finite is called *denumerable*. In this case an onto function $f:N \to X$ gives rise to an "enumeration," $x_1 = f(1), x_2 = f(2), \ldots, x_n = f(n), \ldots$ of the elements of X, so that we may write $X = \{x_1, x_2, \ldots, x_n, \ldots\}$. Since we have not required the function f to be one-one, a given element $x \in X$ may occur more than once in this enumeration. It is easy to see, however, that by deleting all but the first occurrence of any element $x \in X$ and reducing the succeeding subscripts accordingly, it is possible to obtain an enumeration of X in which each element occurs one and only one time.

There are several facts about countability that are of general interest. As a simple consequence of Definition 6.1 we obtain:

Corollary 6.2 Let X and Y be non-empty sets. If X is countable and there is an onto function $g:X \to Y$, then Y is countable.

Proof. Since X is countable, there is an onto function $f:N \to X$, N the set of positive integers. The composite function $gf:N \to Y$ is onto and hence Y is countable.

Corollary 6.3 A non-empty subset of a countable set is countable.

Proof. Let $A \subset X$, X countable, A non-empty. We may define an onto function $g:X \to A$ by setting $g(a) = a$ for $a \in A$ and defining g arbitrarily for points $x \notin A$.

The set N of positive integers is countable, since the identity function $i:N \to N$ is onto. On the other hand, the collection 2^N of subsets of N is not countable, since for an arbitrary set X there is no onto function $f:X \to 2^X$ [see Exer-

cise 1]. A set that is not countable is called *uncountable*. Another example of an uncountable set is the set R of real numbers [see Exercise 2]. Surprisingly, $N \times N$ is a countable set.

Theorem 6.4 Let N be the set of positive integers. Then $N \times N$ is countable.

Proof. The elements of $N \times N$ may be arrayed in the form of the infinite matrix of Figure 33. We may arrange these elements

$$\begin{matrix} (1,1) & (1,2) & (1,3) & \cdots & (1,n) & \cdots \\ (2,1) & (2,2) & (2,3) & \cdots & (2,n) & \cdots \\ (3,1) & (3,2) & (3,3) & \cdots & (3,n) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (m,1) & (m,2) & (m,3) & \cdots & (m,n) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{matrix}$$

Figure 33

in the form of a sequence, $x_1 = f(1)$, $x_2 = f(2), \ldots, x_k = f(k), \ldots$, by setting $x_1 = (1, 1)$, $x_2 = (2, 1)$, $x_3 = (1, 2)$, $x_4 = (3, 1), \ldots$; that is, having exhausted the entries on the diagonal of this matrix from $(p, 1)$ to $(1, p)$ we proceed to enumerate the entries on the diagonal from $(p + 1, 1)$ to $(1, p + 1)$. To explicitly define the onto function $f: N \to N \times N$ we note that there are $\frac{p(p+1)}{2}$ entries on or above the diagonal from $(p, 1)$ to $(1, p)$, hence if $1 \leq j \leq p + 1$ we are setting

$$x_{\frac{p^2+p}{2}+j} = f\left(\frac{p^2+p}{2} + j\right) = (p - j + 2, j).$$

As a direct consequence of Theorem 6.4 and Corollary 6.2 one obtains the result that the set Q^+ of positive rational numbers is countable, for the function $h: N \times N \to Q^+$ defined by $h(r, s) = \frac{r}{s}$, $(r, s) \in N \times N$ is onto.

Corollary 6.5 Let $X_1, X_2, \ldots, X_n \ldots$, be a sequence of sets, each of which is countable. Then

$$\bigcup_{i=1}^{\infty} X_i$$

is a countable set.

Proof. Since each X_i is countable there is an onto function $f_i: N \to X_i$, $i = 1, 2, \ldots, n, \ldots$. We define a function $F: N \times N \to \bigcup_{i=1}^{\infty} X_i$ by setting

$$F(i, j) = f_i(j), \quad (i, j) \in N \times N.$$

F is onto, for if $x \in \bigcup_{i=1}^{\infty} X_i$, $x \in X_i$ for some i, whence $x = f_i(j) = F(i, j)$ for some $(i, j) \in N \times N$. But $N \times N$ is countable and therefore $\bigcup_{i=1}^{\infty} X_i$ is countable.

A more direct proof of Corollary 6.5 can be given by utilizing the matrix of Figure 33 to display the elements of $\bigcup_{i=1}^{\infty} X_i$, entering the element $x_j^i = f_i(j) = F(i, j)$ in the i^{th} row and j^{th} column. One then enumerates the elements of $\bigcup_{i=1}^{\infty} X_i$ in accordance with the scheme adopted in the proof of Theorem 6.4. Since any countable collection of sets may be arranged in the form of a finite or infinite sequence of sets, Corollary 6.5 states that, if X is the union of a countable collection of sets, each of which is countable, then X is countable.

In view of the fact that the set Q^+ of positive rational numbers is countable, the set Q^- of negative rational numbers is also countable. Consequently, the set Q of all rational

numbers is countable. Using Corollary 6.5 we may then assert that the collection B of all open intervals on the real line of the form $S(p; q)$, $q > 0$, with p and q rational, is also a countable set, for it is a countable union of sets each of which is countable. This fact may be used to prove that there is a countable basis for the open sets on the real line.

Let us now return to our discussion of the relation between compactness and the Bolzano-Weierstrass property. The Bolzano-Weierstrass property implies that each countable covering has a finite subcovering.

Theorem 6.6 Let E be a subspace of a topological space X with the property that each infinite subset of E has a point of accumulation in E. Then every countable open covering of E has a finite subcovering.

Proof. We may assume that a countable open covering of E is given in the form of a sequence $O_1, O_2, \ldots, O_n, \ldots$ of open subsets of X such that $E \subset \bigcup_{n=1}^{\infty} O_n$. Suppose that no finite subcollection covers E. Then for each integer k, the open set $O_k^* = \bigcup_{n=1}^{k} O_n$ does not cover E. Hence for each k there is a point $x_k \in E$ such that $x_k \notin O_k^*$. The subset $A = \{x_1, x_2, \ldots, x_k, \ldots\}$ of E must be infinite. Let $x \in E$ be a point of accumulation of A. Since $x \in E$, $x \in O_p$ for some index p. O_p is a neighborhood of x and therefore infinitely many of the points of A belong to O_p. In particular, for some $k > p$ we would have $x_k \in O_p \subset O_p^* \subset O_k^*$, contradicting the choice of x_k. Therefore there must be a finite subcollection of the open sets $O_1, O_2, \ldots, O_n, \ldots$ that covers E.

If a topological space X is such that every open covering has a countable subcovering, by virtue of Theorem 6.6, the Bolzano-Weierstrass property implies compactness. A sufficient condition for every open covering to have a countable

subcovering is given by the next theorem, often called Lindelöf's theorem.

Theorem 6.7 Let X be a topological space that has a countable basis for the open sets. Then each open covering $(O_\alpha)_{\alpha \in I}$ has a countable subcovering.

Proof. Let $\mathcal{B} = (B_\beta)_{\beta \in J}$ be a countable basis for the open sets of X. We shall first prove that for each point $x \in X$ and each open set O containing x, there is a basis element B_β such that

$$x \in B_\beta \subset O.$$

For, since $\mathcal{B}$ is a basis for the open sets, O is a union of elements of $\mathcal{B}$, thus $O = \bigcup_{\beta \in J'} B_\beta$ for some subset J' of J. But $x \in O$, hence $x \in B_\beta$ for some $\beta \in J'$, and clearly $B_\beta \subset O$. Now suppose that $(O_\alpha)_{\alpha \in I}$ is an open covering of x. We must find a countable subset $I' \subset I$ such that $(O_\alpha)_{\alpha \in I'}$ is a covering. For each $x \in X$ and each O_α containing x, we choose a B_β such that $x \in B_\beta \subset O_\alpha$. The totality of sets B_β so chosen constitute a countable subfamily $(B_\beta)_{\beta \in J'}$ of the basis $\mathcal{B}$ and this subfamily covers X. Now, for each such B_β with $\beta \in J'$, let us choose a single index $\alpha = f(\beta) \in I$ such that $B_\beta \subset O_\alpha = O_{f(\beta)}$. The totality of sets O_α so chosen constitute a subfamily $(O_\alpha)_{\alpha \in I'} = (O_{f(\beta)})_{\beta \in J'}$, which is also countable and must cover X, for $\bigcup_{\beta \in J'} B_\beta \subset \bigcup_{\alpha \in I'} O_\alpha$.

Corollary 6.8 Let X be a topological space that has a countable basis for the open sets. Then X is compact if and only if X has the Bolzano-Weierstrass property.

Although we shall not give an example of a topological space X that has the Bolzano-Weierstrass property, but is not compact, the preceding discussion has revealed that such a space must be found among those topological spaces which are not metrizable and do not possess a countable basis for the open sets. Those spaces which possess a countable base for the open sets are called *completely separable* or are said to satisfy the *second axiom of countability*.

Exercises

1. Let X be an arbitrary non-empty set and $f: X \to 2^X$ an arbitrary function from X to the subsets of X. Let A be the subset of X consisting of those points $x \in X$ such that $x \notin f(x)$. Prove that there cannot be a point $a \in X$ such that $A = f(a)$. Finally, prove that there is no onto function $f: X \to 2^X$.

2. Let a function $f: N \to [0, 1]$ be given, N the set of positive integers. In the resulting enumeration $x_1 = f(1)$, $x_2 = f(2)$, . . . , of numbers in $[0, 1]$, express each number x_k in decimal notation $x_k = .a_1^k a_2^k \ldots a_n^k \ldots$, a_n^k an integer $0 \leq a_n^k \leq 9$. Construct a real number $y = .y_1y_2 \ldots y_n \ldots$ such that $y_r \neq a_r^r$, $r = 1, 2, \ldots$, thereby obtaining the result that f cannot be onto and consequently the real numbers are not countable.

3. Use the rational density theorem, which states that between any two real numbers there is a rational number, to prove that the collection of open intervals $S(p; q)$, $q > 0$, p, q rational are a basis for the open sets of R and that therefore R satisfies the second axiom of countability.

4. Let X and Y be topological spaces satisfying the second axiom of countability. Prove that $X \times Y$ also satisfies the second axiom of countability and hence R^n does.

5. Let $(A_\alpha)_{\alpha \in I}$ and $(B_\beta)_{\beta \in J}$ be families of subsets of a set X. $(A_\alpha)_{\alpha \in I}$ is called a refinement of $(B_\beta)_{\beta \in J}$ if for each $\alpha \in I$ there is a $\beta \in J$ such that $A_\alpha \subset B_\beta$. Suppose that $(A_\alpha)_{\alpha \in I}$ is a refinement of $(B_\beta)_{\beta \in J}$ and that $(A_\alpha)_{\alpha \in I}$ covers X. Prove that if I is finite there is a finite subcovering of $(B_\beta)_{\beta \in J}$ and if I is countable there is a countable subcovering of $(B_\beta)_{\beta \in J}$.

6. Recall that a subset A of a topological space X is called *dense* in X if $\overline{A} = X$. A topological space X is called *separable* if there is a countable dense subset. Prove that X is separable if X satisfies the second axiom of countability.

7. A topological space X is said to satisfy the *first axiom of countability* if at each point $x \in X$ there is a countable basis for the complete system of neighborhoods at x. Prove that if X satisfies the second axiom of countability then X satisfies the first axiom of countability.

7 Identification Topologies and Spaces

Let $f: X \to Y$ be a function from a set X onto a set Y. Viewing f as a transformation, for each point $b \in Y$ we may think of f as bringing together or "identifying" the various points in $f^{-1}(\{b\})$. As an example, let us see how the circle, S^1, may be viewed as an interval whose end points have been "identified" or "pasted together." Let $F: [0, 1] \to S^1$ be defined by $F(t) = (\cos 2\pi t, \sin 2\pi t)$. For the moment let us refer to the point $s_0 = (1, 0) \in S^1$ as the base point of S^1. We may think of F as a transformation that "wraps" the unit interval $[0, 1]$ around the circle S^1 so that the two end points meet at the base point s_0. F, in effect, "identifies" the two points 0, $1 \in F^{-1}(\{s_0\})$. The particular function F just considered is a continuous function from a compact space X onto a Hausdorff space Y. Of particular interest is the fact that under these circumstances the topology of Y is determined by the function F and the topology of X.

Lemma 7.1 Let $f: X \to Y$ be a continuous mapping of a compact space X onto a Hausdorf space Y. Then a subset B of Y is closed if and only if $f^{-1}(B)$ is a closed subset of X.

Proof. This lemma is a weaker form of Theorem 2.14. First suppose B is closed. Then $f^{-1}(B)$ is closed by the continuity of f. Conversely, if $f^{-1}(B)$ is closed, then $f^{-1}(B)$ is compact. $B = f(f^{-1}(B))$, hence B is compact. Being a compact subset of a Hausdorff space, B is closed.

Corollary 7.2 Let $f: X \to Y$ be a continuous mapping of a compact space X onto a Hausdorff space Y. Then a subset U of Y is open if and only if $f^{-1}(U)$ is an open subset of X.

To see that in our example $F: [0, 1] \to S^1$ of the preceding paragraph, the topology of S^1 is determined by F and the topology of $[0, 1]$, we may think of each point $t \in [0, 1]$ as "covering" the point $F(t) \in S^1$. A subset U of S^1 not containing the base point s_0 is then open in S^1 if and only if it is "covered" by the open set $F^{-1}(U)$, whereas a subset U of S^1 containing s_0 is open in S^1 if and only if it is "covered" by the open set $F^{-1}(U)$ consisting of the union of an open set U_0 containing 0 and an open set U_1 containing 1. A topology that has been determined in this fashion is called an "identification" topology.

Definition 7.3 Let X be a topological space, Y an (untopologized) set and $f: X \to Y$ a function from X *onto* Y. Let a subset B of Y be called closed if and only if $f^{-1}(B)$ is closed. The resulting topology on Y is called the *identification topology induced by* $f: X \to Y$.

We must justify this definition by proving that the subsets of Y that have been called closed could be the closed subsets of a topological space. Since $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(Y) = X$ are closed in X, the empty set and Y are closed subsets of Y. Suppose, for each $\alpha \in I$, B_α is a closed subset of Y; then $f^{-1}(\bigcap_{\alpha \in I} B_\alpha) = \bigcap_{\alpha \in I} f^{-1}(B_\alpha)$ is the intersection of closed subsets and hence closed, so that $\bigcap_{\alpha \in I} B_\alpha$ is closed. Finally, suppose $B_1, B_2, \ldots, B_n$ are closed subsets of Y. Then

$$f^{-1}(B_1 \cup B_2 \cup \ldots \cup B_n) = f^{-1}(B_1) \cup f^{-1}(B_2) \cup \ldots \cup f^{-1}(B_n)$$

is closed and hence $B_1 \cup B_2 \cup \ldots \cup B_n$ is closed.

In the event that $f: X \to Y$ is a continuous mapping of a compact space X onto a Hausdorff space Y, Lemma 7.1 states that the topology of Y must necessarily have been the identification topology induced by f. If a topological space Y has been obtained by defining an identification topology, the following proposition furnishes a criterion for the continuity of a function g defined on Y.

Proposition 7.4 Let X, Y, Z be topological spaces. Let $f: X \to Y$ be onto and let Y have the identification topology induced by f. Then a function $g: Y \to Z$ is continuous if and only if the composite function $gf: X \to Z$ is continuous.

Proof. Clearly, the fact that Y has the identification topology induced by f makes f continuous, for the inverse image, $f^{-1}(B)$, of a closed subset B of Y is closed. Thus if $g: Y \to Z$ is continuous, the composite function $gf: X \to Y$ is continuous. On the other hand, suppose gf is continuous. We must show that g is continuous. If V is a closed subset of Z, then $(gf)^{-1}(V) = f^{-1}(g^{-1}(V))$ is a closed subset of X. Applying Definition 7.3 to the subset $g^{-1}(V)$ of Y we may conclude that $g^{-1}(V)$ is closed and that therefore g is continuous.

Proposition 7.4 states that if the diagram of topological spaces and functions depicted in Figure 34 is commutative and

Figure 34

$$\begin{tikzcd} X \arrow[d, "f"'] \arrow[dr, "h"] & \\ Y \arrow[r, "g"] & Z \end{tikzcd}$$

Y has the identification topology induced by the onto mapping $f: X \to Y$, then g is continuous if and only if h is continuous. We shall use this result to show that the 2-sphere, S^2, may be viewed as a disc D whose circular boundary has been "shrunk" to a point. We shall do so by constructing first a set Y and a function F from D onto Y so that F "collapses" or "identifies" the boundary circle of D to a single point. We then give Y the identification topology and define functions $G: Y \to S^2$ and $H: D \to S^2$ such that the diagram depicted in Figure 35 is commutative, and G is one-one. Finally, continuity of H will imply that G is a homeomorphism (Theorem 2.14).

$$\begin{tikzcd} D \arrow[d, "F"'] \arrow[dr, "H"] & \\ Y \arrow[r, "G"'] & S^2 \end{tikzcd}$$

Figure 35

Let D be the subset of R^2 consisting of all points (x_1, x_2) such that

$$x_1^2 + x_2^2 \leqq 1.$$

In the plane R^2, the boundary of D is the subset of R^2 defined by

$$x_1^2 + x_2^2 = 1,$$

that is, Bdry $D = S^1$. By S^2 we understand the surface consisting of all points $(x, y, z) \in R^3$ such that

$$x^2 + y^2 + z^2 = 1.$$

Let Y be the (untopologized) set consisting of the points of $D - S^1$; that is, the interior points of D, and a new point s^*. We define a function $F\colon D \to Y$ by setting

$$F(x) = x, x \in D - S^1,$$
$$F(x) = x^*, x \in S^1.$$

As a transformation, we may think of F as leaving all the points in the interior of D fixed, but transforming the boundary S^1 of D into a single point s^*. The function $F\colon D \to Y$ is onto, so we may convert the set Y into a topological space by giving Y the identification topology induced by F. This defines the mapping $F\colon D \to Y$ of our diagram.

We next define a continuous function $H\colon D \to S^2$ by placing the center point $(0, 0)$ of D at the north pole $(0, 0, +1)$ of S^2 and "wrapping" D around the sphere by "pulling together" the boundary circle of D to a point at the south pole $(0, 0, -1)$

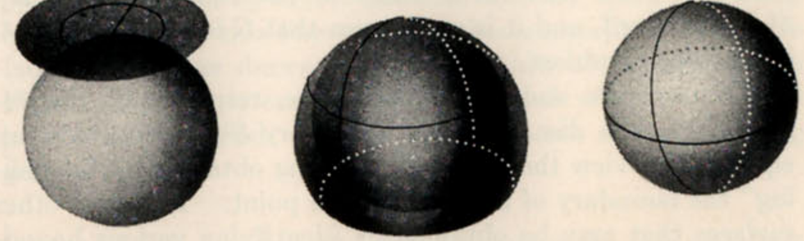


Figure 36

of S^2, [this process is pictured in Figure 36]. Analytically, the function H could be given by taking a point $(x_1, x_2) \in D$, writing it in polar form, $(x_1, x_2) = (\rho \cos t, \rho \sin t)$, and then setting $H(x_1, x_2) = (\sin \pi\rho \cos t, \sin \pi\rho \sin t, \cos \pi\rho)$. In this way the line segment L defined by $x_2 = 0$ or $t = 0$ is carried into the circle of longitude $H(L)$ consisting of the points $(\sin \pi\rho, 0, \cos \pi\rho)$, $-1 \leqq \rho \leqq 1$, whereas the circle Q defined by $x_1^2 + x_2^2 = \frac{1}{2}$ or $\rho = \frac{1}{2}$ is carried into the equator $H(Q)$ consisting of the points $(\cos t, \sin t, 0)$, $0 \leqq t \leqq 2\pi$ and the boundary S^1 of D is carried into the south pole $H(S^1) = (0, 0, -1)$. We have now defined the function $H\colon D \to S^2$, which we may assert to be continuous either on intuitive grounds such as exhibited in Figure 36 or because H may be defined so that the x, y, z coordinates of $H(x_1, x_2)$ are continuous functions.

The last function we must define is the function $G\colon Y \to S^2$. We do so by setting

$$G(x) = H(x), x \in D - S^1,$$
$$G(x^*) = (0, 0, -1).$$

For $x \in D - S^1$, $G(F(x)) = G(x) = H(x)$, whereas for $x \in S^1$, $G(F(x)) = G(s^*) = (0, 0, -1) = H(x)$. Thus $GF = H$ and we have constructed a commutative diagram. Since H is continuous, by Proposition 7.4, G is continuous. Now, Y is com-

pact [Y is the continuous image of the compact space D], S^2 is Hausdorff, and it is easily seen that G is one-one, thus G is a homeomorphism.

A rectangle and its boundary are, respectively, homeomorphic to the disc D and its boundary S^1, so that we may equally well view the sphere S^2 as being obtained by "shrinking" the boundary of a rectangle to a point. There are other surfaces that may be obtained by identifying various boundary points of a rectangle. Perhaps the simplest of these surfaces is a cylinder. If we take a rectangle with four corner vertices A, B, B', A' [see Figure 37a] and identify the edge AB with the edge $A'B'$ in such a way that A is identified with A' and B with B', then we obtain a surface that is homeomorphic to the cylinder in Figure 37b. We may equally well picture the cylinder as being the topological space obtained by replacing both A and A' by a new point A^*, both B and B' by a new point B^*, and similarly any pair of corresponding points C and C' on the respective edges AB and $A'B'$ is replaced by a new point C^* as indicated in Figure 37c. Furthermore, a neighborhood of this new point C^* would contain the interior of the small semi-circles drawn in Figure 37c. It is interesting to note that if in this figure we join C^* to itself by the

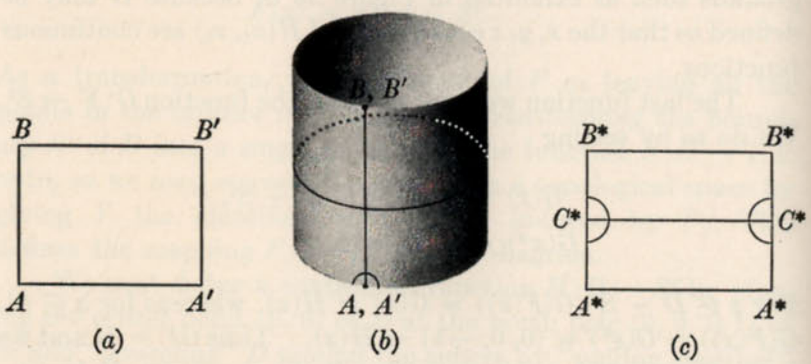


Figure 37

path represented by the horizontal line, the space consisting of the points of this line would be homeomorphic to a circle [such as the one drawn about the middle of the cylinder in Figure 37b], for it consists of an interval whose end points have been identified. This is a special case of the following general result.

Lemma 7.5 Let X and Y be topological spaces, let $f: X \to Y$ be a continuous function that is onto, and let Y have the identification topology induced by f. If $B \subset Y$ is such that $A = f^{-1}(B)$ is closed, then the subspace B of Y has the identification topology induced by the restriction $f \mid A: A \to B$.

Proof. We must show that a subset F of B is closed in B if and only if $(f \mid A)^{-1}(F)$ is closed in A. The restriction $f \mid A$ of the continuous function f to $A = f^{-1}(B)$ is continuous, so that if F is closed in B, then $(f \mid A)^{-1}(F)$ is closed in A. Conversely, suppose that $(f \mid A)^{-1}(F)$ is closed in A. Then, since A is closed in X, $(f \mid A)^{-1}(F)$ is closed in X. If we prove that $(f \mid A)^{-1}(F) = f^{-1}(F)$, it will follow that F is closed in Y and consequently in B, for Y has the identification topology and therefore $f^{-1}(F)$ closed in X implies F closed in Y. It remains to prove $(f \mid A)^{-1}(F) = f^{-1}(F)$. Suppose that $x \in f^{-1}(F)$. To show that $x \in (f \mid A)^{-1}(F)$ we must show that $x \in A$ and $(f \mid A)(x) \in F$. But if $x \in f^{-1}(F)$, then $f(x) \in F \subset B$, whence $x \in f^{-1}(B) = A$. Thus x is in the domain of $f \mid A$ and $(f \mid A)(x) = f(x) \in F$, hence $x \in f^{-1}(F)$ implies that $x \in (f \mid A)^{-1}(F)$. Conversely, if $x \in (f \mid A)^{-1}(F)$, then $(f \mid A)(x) \in F$. Now $(f \mid A)(x) = f(x)$, thus $f(x) \in F$ and $x \in f^{-1}(F)$. It follows that $(f \mid A)^{-1}(F) = f^{-1}(F)$, and the proof is complete.

Another surface that may be obtained by identifying some of the boundary points of a rectangle is a surface called the Möbius strip or band. Starting again with the rectangle whose vertices we shall now label in the order A, B, A', B' [see Figure 38a], we identify the edge AB with the edge $B'A'$ by first giving the rectangular strip a 180 degree twist, so that the vertices A and A' coincide and the vertices B and B' coincide [Figure 38b]. A topologically equivalent space is in-

dicated in Figure 22c, where corresponding or identified pairs of points such as A, A' have been replaced by a single new point A^*. The fact that we intend to identify the two edges AB and $A'B'$ of Figure 38a with a twist is often indicated

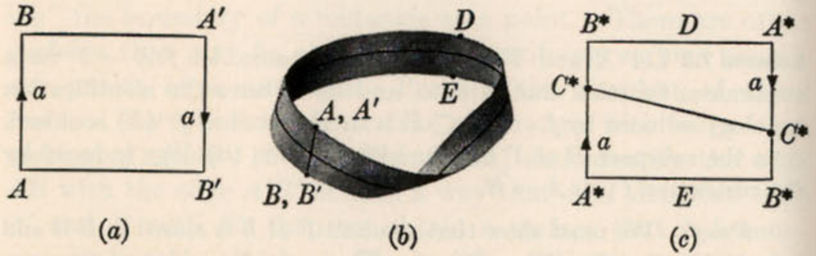


Figure 38

by labelling the edges with the same letter, such as "a," and then placing arrowheads on these edges in such a position that the resulting identification matches up or superimposes the two arrowheads. The Möbius strip has many curious properties. The oblique line in Figure 38c joining C^* to itself is homeomorphic to a circle. The upper horizontal line running from B^* through D to A^* is homeomorphic to an interval. However, if on the Möbius strip we trace out the curve from B^* through D to A^* and continue on [along the lower horizontal line of Figure 38c] through E back to B^* we trace out an interval with its end-points identified, that is, a circle. Thus the Möbius strip is a surface whose bounding curve is a circle. Other interesting properties may be deduced from the representation in Figure 38c. For example, if the Möbius strip is cut down its center, the resulting surface will not be disconnected for we may still connect a point of the upper half rectangle in Figure 38c to a point of the lower half rectangle by joining both of them to the bounding curve $B^*DA^*EB^*$.

So far we have considered only surfaces resulting from the identification of a pair of edges of a rectangle. If we identify the edges of a rectangle according to the scheme indicated in Figure 39a, the resulting topological space is called a *torus*.

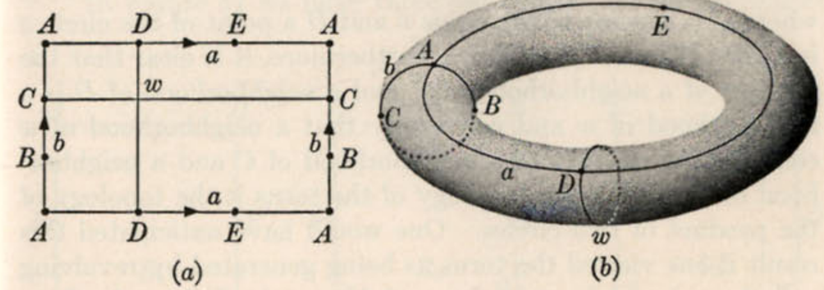


Figure 39

A torus is topologically the surface of a donut or a rubber tire, as indicated in Figure 39b. We may view the torus as being obtained in two steps. First, we identify the two opposite edges labelled a of the rectangle to obtain a cylinder, and second, we identify the two resulting circular edges (labelled b) of the cylinder to obtain the torus. The justification for breaking the identification up into two steps is contained in the following proposition.

Proposition 7.6 Let X, Y, Z be topological spaces, let $f:X \to Y$ and $g:Y \to Z$ be continuous and onto. If Y has the identification topology induced by $f:X \to Y$ and Z has the identification topology induced by $g:Y \to Z$, then Z has the identification topology induced by $gf:X \to Z$.

Proof. Clearly, if F is a closed subset of Z, then $(gf)^{-1}(F)$ is a closed subset of X, for gf is continuous. Conversely, suppose $(gf)^{-1}(F) = f^{-1}(g^{-1}(F))$ is a closed subset of X. Since Y has the

identification topology induced by $f: X \to Y$, $g^{-1}(F)$ is a closed subset of Y. Similarly, since Z has the identification topology induced by $g: Y \to Z$, $g^{-1}(F)$ closed in Y implies that F is closed in Z. Thus F is closed if and only if $(gf)^{-1}(F)$ is closed; that is, Z has the identification topology induced by $gf: X \to Z$.

Topologically, the torus is the product of two circles. An arbitrary point w of the torus may be written as $w = (C, D)$, where C is a point of the circle b and D a point of the circle a in either Figure 39a or 39b. Furthermore, it is clear that the product of a neighborhood of C and a neighborhood of D is a neighborhood of w and conversely that a neighborhood of w contains the product of a neighborhood of C and a neighborhood of D. Thus the topology of the torus is the topology of the product of two circles. One would have anticipated this result if one viewed the torus as being generated by revolving a circle such as b in a circular path by moving it in such a way as to always have the point labelled A in contact with the circle labelled a.

There are two other surfaces resulting from the identification of opposite pairs of edges of a rectangle. One of these surfaces is called a *Klein bottle*. The Klein bottle may be obtained by first identifying the edges labelled a in Figure 40a in the prescribed manner to obtain a cylinder, and then identifying the two circles labelled b in either Figure 40a or 40b,

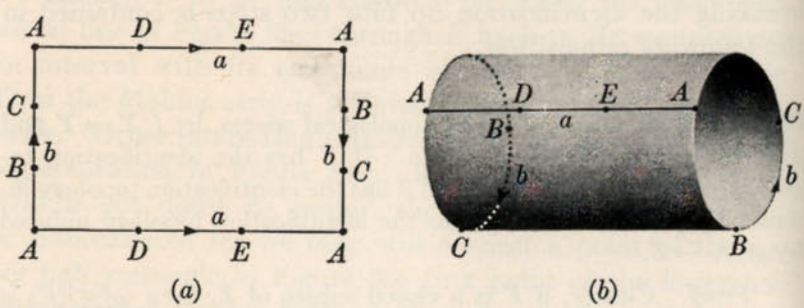


Figure 40

not, however, in the manner of Figure 39 to obtain a torus, but with a "twist." Unfortunately, at least from the point of view of our visualization of the Klein bottle, there is no way to identify these two circular edges of the cylinder of Figure 40b without forcing the surface of the Klein bottle to intersect or pass through itself. For this reason, it is helpful to construct the Klein bottle in several pieces.

In Figure 41 we have three rectangles. If the rectangles

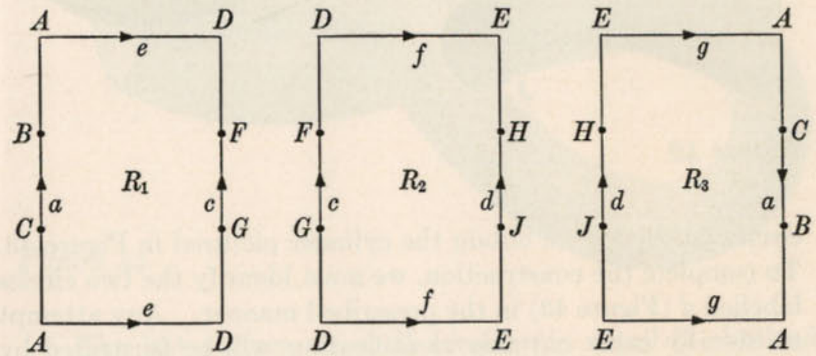


Figure 41

R_1 and R_2 are joined along the edge labelled c and the rectangles R_2 and R_3 are joined along the edge labelled d, we obtain the rectangle and identifications of Figure 40, so that Figure 41 also represents the Klein bottle. If, in these three rectangles, we first identify the pairs of edges labelled e, f, and g respectively, we obtain three cylinders that are homeomorphic to the three corresponding cylindrical surfaces of Figure 42, also labelled R_1, R_2, R_3. To construct the Klein bottle we need only identify these three cylinders along the pairs of circular edges labelled a, c, and d, respectively. We may join the cylinders R_1 and R_3 along the circles labelled a, so that R_3 lies inside R_1. If we then join R_1 and R_2 along the

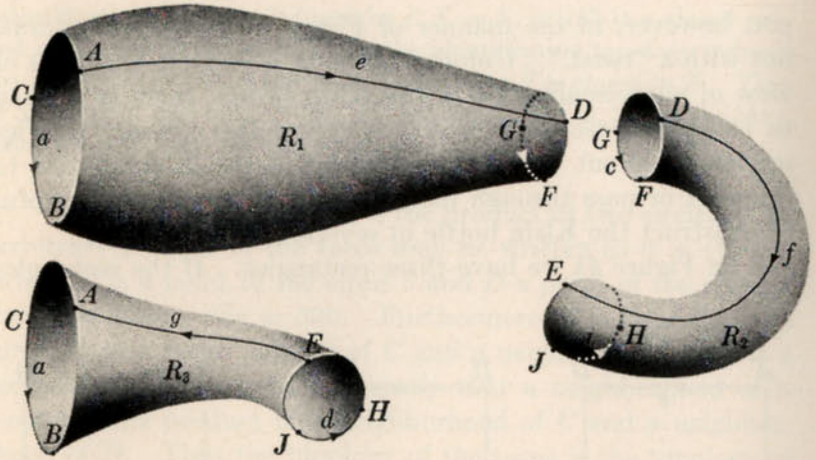


Figure 42

circles labelled c, we obtain the cylinder pictured in Figure 43. To complete the construction, we must identify the two circles labelled d (Figure 43) in the prescribed manner. Any attempt to literally carry out this identification will be frustrated by our inability to pass through the surface of the cylinder. We must therefore either be content, as in Figure 43, to indicate

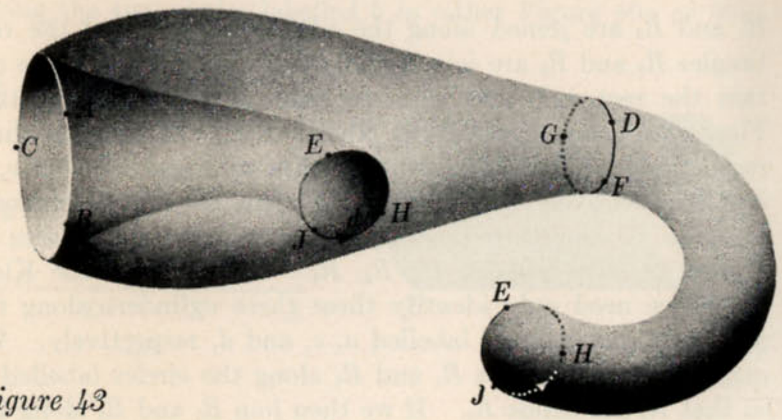


Figure 43

this identification, or adopt the fiction that in Figure 44 the Klein bottle does not intersect itself along the circle d, but that each point along d is to represent at the same time two points of the Klein bottle.

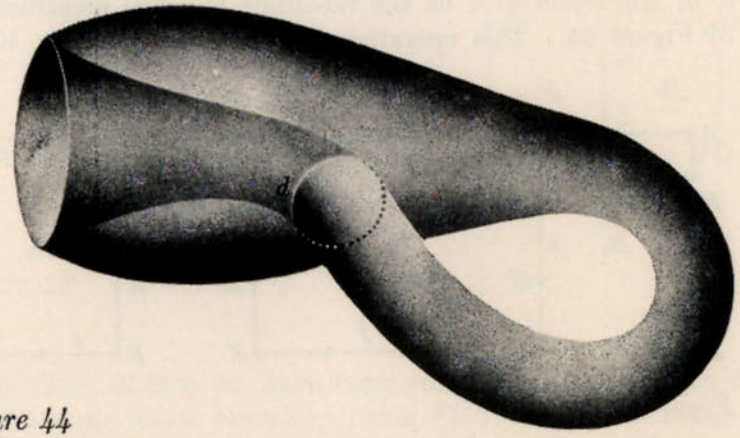


Figure 44

The last surface we shall consider in detail is obtained by identifying both of the pairs of opposite edges of a rectangle with a "twist." These identifications are indicated in Figure 45. Note that in this figure all the vertices are not

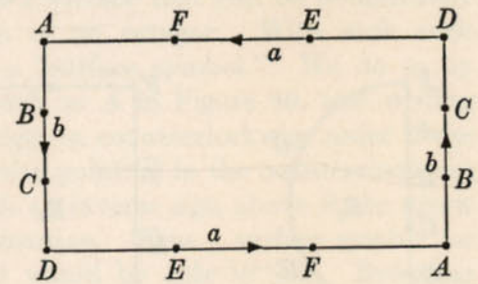


Figure 45

identified with one another, but only diagonally opposite vertices are joined together. In order to relate this surface

to some of the preceding surfaces, we shall adopt the same method as the one used in the examination of the Klein bottle, [one might call this the "cut-and-paste method"]. We first separate the large rectangle into three smaller rectangles R_1, R_2, R_3, which when re-identified along the pairs of edges labelled c, d, will again give us the rectangle and the identifications of Figure 45. This operation is indicated in Figure 46. If

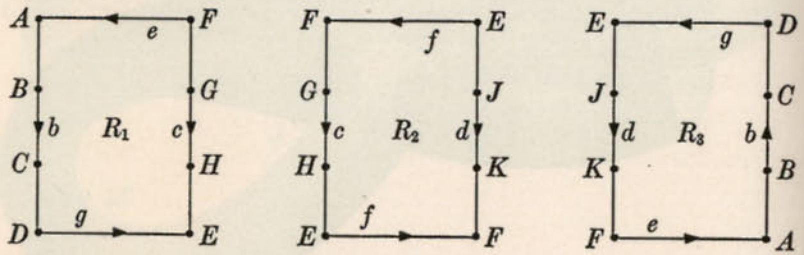


Figure 46

we first join the two edges labelled f in rectangle R_2 we obtain a Möbius strip. Since we are only interested in the topological nature of this surface, we may distort [by homeomorphisms] the two rectangles R_1 and R_3 into the semicircular regions of Figure 47. If we then join the regions R_1 and R_3

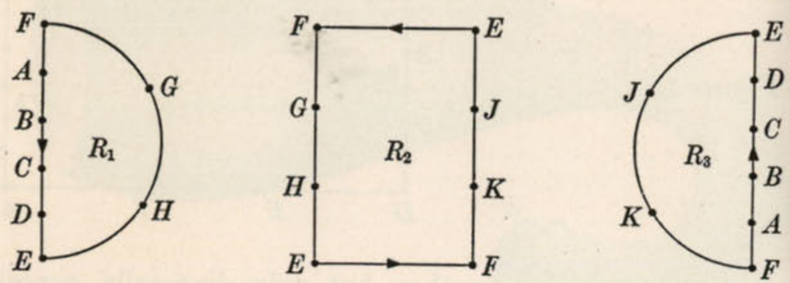


Figure 47

along their common edge $FABCDE$ we obtain the disc and the Möbius strip of Figure 48, with the indicated identifica-

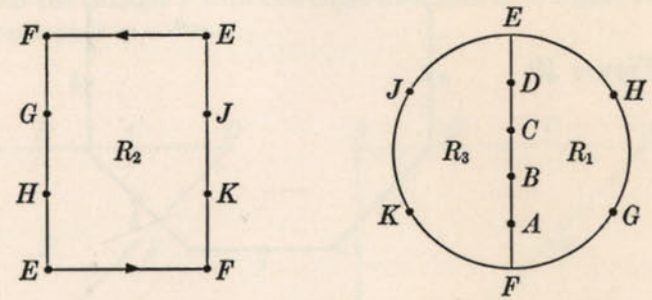


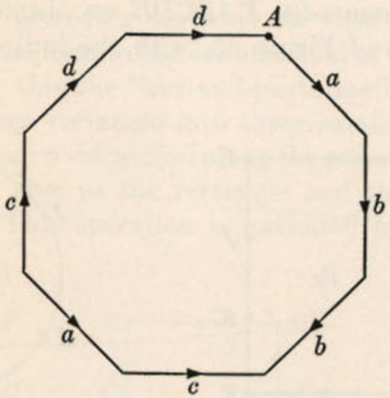
Figure 48

tions. The surface we have been considering is therefore a Möbius strip whose boundary circle $FGHEJKF$ is to be attached to the boundary circle $FGHEJKF$ of a disc. This last surface is called the *projective plane.*

The sphere, torus, Klein bottle, and projective plane are examples of a larger class of surfaces that may be obtained by identifying pairs of edges of a polygon with $2n$ sides. Such surfaces are called closed 2-manifolds. For example, in Figure 49 we have indicated a surface that can be obtained by identifying pairs of sides of an octagon. With each such figure we may associate a "surface symbol." We do so by starting at any vertex, such as A in Figure 49, and writing down the labels of the edges in counterclockwise order if the arrow along that edge is also pointing in the counterclockwise direction or the label with an inverse sign above if the arrow points in the clockwise direction. Thus a surface symbol for the surface of Figure 49 would be $abbc^{-1}a^{-1}cdd$. Referring back to Figure 39, one can see that a surface symbol for the torus is $ab^{-1}a^{-1}b$.

By the "cut-and-paste" method one can show that each

Figure 49



2-manifold is homeomorphic to a 2-manifold whose surface symbol is of one of the following four forms: $abb^{-1}a^{-1}$; $a_1b_1a_1^{-1}b_1^{-1} \ldots a_pb_pa_p^{-1}b_p^{-1}$, $p \geqq 1$; $abab$; $a_1a_1 \ldots a_qa_q$, $q > 1$. The first form indicates that the surface is homeomorphic to a sphere. The second form includes the surface symbol of a torus and in general indicates that the surface is homeomorphic to a sphere with p handles. These two classes of surface are called *orientable*. They can all be constructed in three-dimensional Euclidean space and are two-sided. The third form indicates that the surface is homeomorphic to the projective plane. We have seen that the projective plane is a disc to whose circular boundary has been attached a Möbius strip. One may think of the disc as constituting the portion of the surface of a sphere obtained by removing a circular region. Attaching a Môbius strip to the circular boundary of this region is called attaching "a crosscap." Thus the projective plane is called "a sphere with crosscap." In the same manner, the second form consists of all surfaces obtained by attaching q Möbius strips or crosscaps to a sphere with q circular regions removed.

The reader interested in a detailed treatment of 2-manifolds will find an excellent presentation in either of the books by Cairns or Lefschetz previously mentioned.

Exercises

1. Prove that the triangle T with two edges identified as in Figure 50 is homeomorphic to a disc.

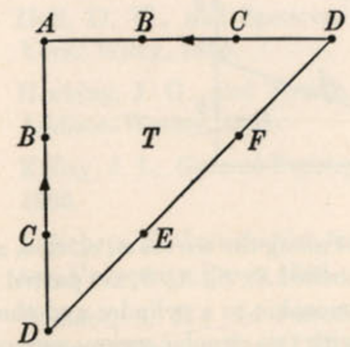


Figure 50

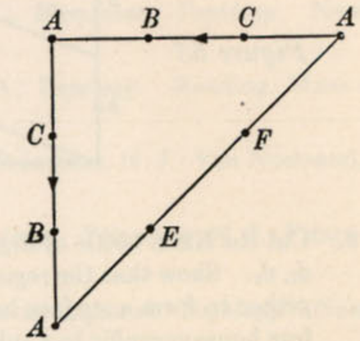
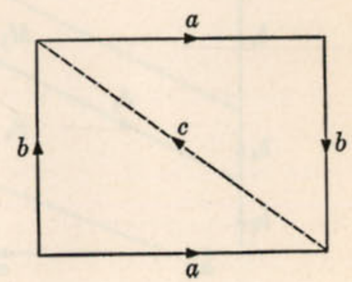


Figure 51

2. Prove that the triangle S with two edges identified as in Figure 51 is a Möbius strip.

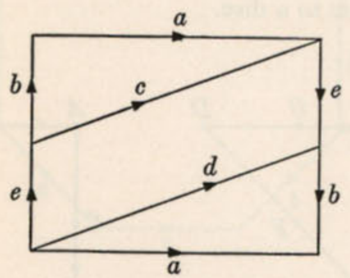
3. Prove that the Klein bottle is homeomorphic to a surface with surface symbol $a_1a_1a_2a_2$ by cutting the rectangle of Figure 52 along the diagonal c and pasting the resulting triangles along their common edge b.

Figure 52



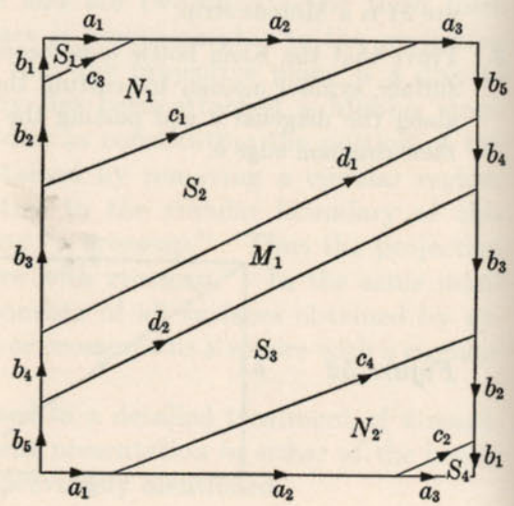
4. Show that if the Klein bottle of Figure 53 is cut along the curves c and d the result is two Möbius strips and that therefore the Klein bottle is two Möbius strips joined along their circular boundaries.

Figure 53



5. Cut the Klein bottle of Figure 54 along the curves c_1, c_2, c_3, c_4 and d_1, d_2. Show that the regions labelled S_1, S_2, S_3, S_4 are pasted together to form a surface homeomorphic to a cylinder and therefore homeomorphic to a sphere with two circular regions removed whose boundaries are the circles d_1d_2 and $c_1c_2c_3c_4$ respectively. Show that the region labelled M_1 is a Möbius strip whose boundary is d_1d_2 and that the regions labelled N_1 and N_2 form a second Möbius strip whose boundary is $c_1c_2c_3c_4$.

Figure 54



Bibliography

This list has been restricted to introductory texts in English. For further titles see the references in Hocking and Young.

1. Cairns, S. S., *Introductory Topology.* New York: Ronald Press, 1961.
2. Hall, D. W., and Spencer, G. L., *Elementary Topology.* New York: Wiley, 1955.
3. Hocking, J. G., and Young, G. S., *Topology.* Reading, Mass.: Addison-Wesley, 1961.
4. Kelley, J. L., *General Topology.* Princeton, N. J.: Van Nostrand, 1955.
5. Lefschetz, S., *Introduction to Topology.* Princeton, N. J.: Princeton University Press, 1942.
6. Wallace, A. H., *An Introduction to Algebraic Topology.* New York: Pergamon Press, 1957.

Special Symbols

$\in$	belongs to, is a member of, in, 4
$\notin$	does not belong to, is not a member of, 4
$\{A_1, A_2, \ldots, A_n\}$	the set whose members are $A_1, A_2, \ldots, A_n$, 4
$\emptyset$	null set or empty set, 5
$\subset$	is contained in, is a subset of, 5
$\supset$	contains, 5
$[a, b]$	closed interval, all real numbers x such that $a \leqq x \leqq b$, 5
(a, b)	open interval, all real numbers x such that $a < x < b$, 5
2^A	set of all subsets of A, 6
$\cap$	intersection, 7
$\cup$	union, 7
$C_S(A)$	complement of A in S, 8
$S - A$	complement of A in S, 8
$C(A)$	complement of A, 8
$(A_\alpha)_{\alpha \in I}$	indexed family of subsets, for each α in the indexing set I, A_α is a subset of a set S, 10
$\bigcup_{\alpha \in I} A_\alpha$	union over α of A_α, all x such that $x \in A_\beta$ for at least one $\beta \in I$, 10
$\bigcap_{\alpha \in I} A_\alpha$	intersection over α of A_α, all x such that $x \in A_\beta$ for each $\beta \in I$, 10
$\bigcup_{i=1}^{n}$	all x such that $x \in A_i$ for at least one integer i, $1 \leqq i \leqq n$, 12
$\bigcap_{i=1}^{n} A_i$	all x such that $x \in A_i$ for each integer i, $1 \leqq i \leqq n$, 12
(x, y)	ordered pair, 13
$A \times B$	Cartesian product of A and B, all ordered pairs (x, y) such that $x \in A$ and $y \in B$, 13
$\prod_{i=1}^{n} A_i$	direct product, all sequences $(a_1, a_2, \ldots, a_n)$ such that $a_i \in A_i$, $1 \leqq i \leqq n$, 13
$\prod_{i=1}^{\infty} A_i$	all infinite sequences $(a_1, a_2, \ldots)$ such that $a_i \in A_i$ for each positive integer i, 14

R^n	Euclidean n-space, all $x = (x_1, x_2, \ldots, x_n)$, x_i a real number, 14
A^n	all $a = (a_1, a_2, \ldots, a_n)$, $a_i \in A$, 14
$f{:}A \to B$	a function f from A to B, 16
$A \xrightarrow{f} B$	a function f from A to B, 16
$f(X)$	image of X, all y such that $y = f(x)$ for some $x \in X$, 16
$f^{-1}(Y)$	inverse image of Y, all x such that $f(x) \in Y$, 16
gf	composition of f and g, the function which associates to each element a the element $g(f(a))$, 20
$f \mid A$	f restricted to A, the function which associates to each x in the subset A of the domain of f the element $f(x)$, 29
(X, d)	metric space whose underlying set is X and whose distance function is d, 33
(R, d)	metric space consisting of the set of real numbers and distance d given by $d(a, b) = \lvert a - b \rvert$, 34
(R^n, d)	metric space consisting of Euclidean n-space R^n and distance d given by $d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \underset{1 \leqq i \leqq n}{\text{maximum}} \{\lvert x_i - y_i \rvert\}$, 35
(R^n, d')	metric space consisting of Euclidean n-space R^n and distance d' given by $d'((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, 36
$f{:}(X, d) \to (X, d')$	a function f from the metric space (X, d) to the metric space (X, d'), 42
$S(a; \delta)$	open sphere about a of radius δ, all x such that $d(x, a) < \delta$, 45
$\lim_n a_n = a$	a is the limit of the sequence $a_1, a_2, \ldots$, 56
g.l.b.	greatest lower bound, 62
$d(a, A)$	distance between a and A, g.l.b. of the set of numbers $d(a, x)$ for $x \in A$, 63
S^n	n-sphere, all $x = (x_1, x_2, \ldots, x_{n+1}) \in R^{n+1}$ such that $x_1^2 + x_2^2 + \ldots + x_{n+1}^2 = 1$, 72
I^n	unit n-cube, all $x = (x_1, x_2, \ldots, x_n) \in R^n$ such that $0 \leqq x_i \leqq 1$ for $i = 1, 2, \ldots, n$, 72

$(X, \mathfrak{I})$	topological space with underlying set X and topology $\mathfrak{I}$, 84
$\bar{A}$	closure of A, all x such that $N \cap A \neq \emptyset$ for each neighborhood N of x, 96
Int (A)	interior of A, all x such that A is a neighborhood of x, 100
Bdry (A)	boundary of A, $\bar{A} \cap \overline{C(A)}$, 101
$(a, b]$	all real numbers x such that $a < x \leqq b$, 123
$[a, b)$	all real numbers x such that $a \leqq x < b$, 123
$(-\infty, a)$	all real numbers x such that $x < a$, 123
$(-\infty, a]$	all real numbers x such that $x \leqq a$, 123
$(a, +\infty)$	all real numbers x such that $x > a$, 123
$[a, +\infty)$	all real numbers x such that $x \geqq a$, 123
Cmp(a)	component of a, largest connected set containing a, 135
ACmp(a)	arc component of a, all y which can be connected to x by a path, 143
$\cong$	homotopic, 150
$[\![a]\!]$	equivalence set of a, 151
$[\![e_z]\!]$	homotopy class of the constant path e_z at z, 153
I	unit interval $[0, 1]$, 147
ε_L	Lebesque number of an open covering $(O_\alpha)_{\alpha \in I}$ of a compactum, 182

Index

DOCUMENTS ON MODERN PHYSICS

A new series of monographs which will include selected reviews, lectures, conference proceedings and important collections of papers on branches of physics of special current interest. The books will be of particular interest to senior undergraduate and graduate students and to professional, experimental and theoretical physicists.

already available

Theory of Crystal Dislocations
A. H. COTTRELL

Lectures on Magnetoionic Theory
K. G. BUDDEN

Solar Plasma, Geomagnetism and Aurora
S. CHAPMAN

Experimental Relativity
R. H. DICKE

in preparation

The Dynamics of Conduction Electrons
A. B. PIPPARD

Dynamical Theory of Groups and Fields
B. S. DEWITT

Geometrodynamics and the Issue of the Final State
J. A. WHEELER

Published by Blackie & Son Ltd.

BOOKS ON MATHEMATICS

THEORY AND APPLICATION OF INFINITE SERIES

KONRAD KNOPP

Translated from the second German edition and revised in accordance with the fourth by R. C. H. Young, Ph.D.

Part I. Real Numbers and Sequences.
Part II. Foundations of the Theory of Infinite Series.
Part III. Development of the Theory.

THE ABSOLUTE DIFFERENTIAL CALCULUS

TULLIO LEVI-CIVITA

Authorized translation by M. Long

A detailed exposition of the calculus of tensors by an eminent author who has been more closely associated with the development of the subject than any other living mathematician.

DIFFERENTIATION AND INTEGRATION

H. A. THURSTON, M.A., PH.D.

A first course in the calculus which differs from previous English texts in that the author believes that the time has come for some steps forward. His approach is forthright and stimulating.

DIFFERENTIAL EQUATIONS

F. G. TRICOMI

Translated from the Italian third edition by Elizabeth A. McHarg, PH.D

A teaching book designed to give the student a clear idea of the problems and methods of the theory of differential equations to Honours degree level.

PUBLISHED BY BLACKIE & SON LTD.

INTRODUCTION TO TOPOLOGY

MENDELSON

BLACKIE